

CURRENT ISSUE

Table of Contents

Volume 480 Number 7375 pp5-144



- THIS WEEK
 - Editorials
 - World View
 - Research Highlights

- Seven Days

- NEWS IN FOCUS

- News
- Features

•

- COMMENT

- Comment
- Books and Arts
- Correspondence
- Obituary

- CAREERS

- Feature
- Q&A

- Futures

•

- SPECIALS

- Technology Feature

- RESEARCH

- Brief Communications Arising
- News & Views
- Review
- Articles
- Letters
- Corrigendum

[About the cover](#)

- [Previous issue](#)

THIS WEEK

EDITORIALS

Top

- **Half way there**

Spain, Italy and Greece all have new governments and new research laws. Despite the pressures of economic austerity, investing in science now could bring disproportionate benefits.

- **False economy**

The Danish government's plan to axe technology assessment is ill-conceived.

- **A poor sequel**

Muted media response to the release of more climate e-mails shows science's strength.

WORLD VIEW

Top

- **Psychology must learn a lesson from fraud case**

Sharing data could help to avert scandals like the Diederik Stapel revelations, and improve the quality of research, says Jelte M. Wicherts.

RESEARCH HIGHLIGHTS

Top

- **Physics: Thunder but no lightning**
- **Medicine: Double radiation shield**
- **Evolution: Big bites help bats diversify**
- **Technology: Low-power magnetic switch**
- **Neuroscience: Sleep calms the emotions**
- **Palaeoanthropology: Diet sculpts human jaws**
- **Organic chemistry: Random mix reveals reaction**
- **Molecular Medicine: Liver disease target**

- **Zoology: Spiders' chemical deterrent**

• COMMUNITY CHOICE

Materials science: Stretchy graphene transistors

SEVEN DAYS

Top

- **Seven days: 25 November–1 December 2011**

The week in science: anti-HIV gel fails Africa trial; funding crunch for Global Fund to Fight AIDS, Tuberculosis and Malaria; and NASA's Curiosity rover bound for Mars.

NEWS IN FOCUS



- **Embattled scientist in theft probe**

Sacked virologist faces lawsuit from Nevada institute.

- Ewen Callaway

○



- **Australia's marine plans questioned**

Ocean reserves offer inadequate protection, critics say.

- Daniel Cressey

○



- **Scientists, meet capitalists**

US agencies and scientific societies aim to create jobs by teaching researchers how to be entrepreneurs.

- Eugenie Samuel Reich

○



- **Blockbuster drug bows out**

Pharmaceutical industry anxiously struggles to retool as Lipitor patent expires.

- Heidi Ledford

○



- **Outcry over EU budget plan**

Costly Earth-monitoring and fusion-energy projects could be stripped from main budget.

- Declan Butler

○

- **Correction**

FEATURES

Top



- **Fighting for the forest: The roadless warrior**

To save the Amazon, Bruce Babbitt wants to isolate islands of oil and gas production amid a sea of trees.

- Jeff Tollefson



- **Cell biology: The new cell anatomy**

A menagerie of intriguing cell structures, some long-neglected and others newly discovered, is keeping biologists glued to their microscopes.

- Roberta Kwok

COMMENT

- **Antarctica: Turning the world upside down**

Research, not pole-bagging, was the lasting achievement of Antarctic exploration 100 years ago, says Edward J. Larson.

- **Climate change: High risk of permafrost thaw**

Northern soils will release huge amounts of carbon in a warmer world, say Edward A. G. Schuur, Benjamin Abbott and the Permafrost Carbon Network.

- **The new cell biology: Beyond HeLa cells**

To find out what distinguishes one cell type from another, cell biologists must renounce popular cell lines, argue Anthony H. Hyman and Kai Simons.

BOOKS AND ARTS

Top

- **Physics: Atomic secrets**

Mystery lingers round the sudden defection of cold-war physicist Bruno Pontecorvo, finds Sharon Weinberger.

- Review of *The Pontecorvo Affair: A Cold War Defection and Nuclear Physics*

Simone Turchetti

- **Museums: Stripped assets**

Paolo Mazzarello argues that the disposal of collections requires clear consultation with the public.

- **Books in brief**

- **Chemistry: An elemental heroine**

An opera on the astonishing life of Marie Curie enthalls Stefan Michalowski and Georgia Smith.

- Review of *Madame Curie*

Elżbieta Sikora & Agata Miklaszewska

CORRESPONDENCE

Top

- **Development models: Arab science to turn spring into summer**

- M. Dahmani Fathallah

- **Double truth: Climate results for public vetting**

- Mathis Hampel, Martin Mahony & Mike Hulme

- **Cloning technology: Control the bonanza for research eggs**

- Marcy Darnovsky, Susan Berke Fogel & Judy Norsigian

- **Food security: A role for Europe**

- Tim Benton, Tibor Hartel & Josef Settele

- **Food security: Close crop yield gap**

- Robert Finger

OBITUARY

Top

- **John McCarthy (1927–2011)**

The scientist who set computers on the path to common sense.

- Vladimir Lifschitz

CAREERS

FEATURES

Top

- **Single-cell analysis: Imaging is everything**

Advances in single-cell imaging bring opportunities for physicists, biologists and chemists alike.

- Amy Maxmen

Q&AS

Top

- **Turning point: Jonathan Cirtain**

Former athlete's move into solar physics and telescope design leads to award.

- Virginia Gewin

naturejobs job listings and advertising features

FUTURES

- **Gifts of the Magi**

A career underground.

- Anatoly Belilovsky

SPECIALS

TECHNOLOGY FEATURE

Top

- **Single-cell analysis: The deepest differences**

To understand biological heterogeneity, researchers are learning how to profile the molecular contents of individual cells.

- Charlotte Schubert

RESEARCH

BRIEF COMMUNICATIONS ARISING

Top

- **Fundamental limits of ‘ankylography’ due to dimensional deficiency**
- Haiqing Wei
- **Non-uniqueness and instability of ‘ankylography’**
- Ge Wang, Hengyong Yu, Wenxiang Cong & Alexander Katsevich

NEWS & VIEWS

Top

● FORUM

Forum Optics: Perfect lenses in focus

- Tomáš Tyc & Xiang Zhang

- **Evolution: Big brains explained**

- Richard Potts

See also

- Letter by Navarrete et al.

- **Regenerative Medicine: Organ recital in a dish**

- Karine Rizzoti & Robin Lovell-Badge

See also

- Article by Suga et al.

- **Physiology: Immune cells fuel the fire**

- Andrew J. Whittle & Antonio Vidal-Puig

See also

- Letter by Nguyen et al.

- **Astrophysics: A puzzling γ-ray burst**

- Enrico Costa

See also

- Letter by Campana et al.

See also

- Letter by Thöne et al.

- **Earth science: Redox state of early magmas**

- Bruno Scaillet & Fabrice Gaillard

See also

- Letter by Trail et al.

REVIEW

Top

- **The case against climate regulation via oceanic phytoplankton sulphur emissions**
- P. K. Quinn & T. S. Bates

ARTICLES

Top

- **Self-formation of functional adenohypophysis in three-dimensional culture**

- Hidetaka Suga, Taisuke Kadoshima, Maki Minaguchi, Masatoshi Ohgushi, Mika Soen+ *et al*

See also

- News & Views by Rizzoti & Lovell-Badge

- **Mutations causing syndromic autism define an axis of synaptic pathophysiology**

- Benjamin D. Auerbach, Emily K. Osterweil & Mark F. Bear

LETTERS

Top

- **The unusual gamma-ray burst GRB 101225A explained as a minor body falling onto a neutron star**

- S. Campana, G. Lodato, P. D’Avanzo, N. Panagia, E. M. Rossi + *et al*

See also

- News & Views by Costa

- **The unusual γ-ray burst GRB 101225A from a helium star/neutron star merger at redshift 0.33**

- C. C. Thöne, A. de Ugarte Postigo, C. L. Fryer, K. L. Page, J. Gorosabel + *et al*

See also

- News & Views by Costa

- **Observation of a pairing pseudogap in a two-dimensional Fermi gas**

- Michael Feld, Bernd Fröhlich, Enrico Vogt, Marco Koschorreck & Michael Köhl

- **The oxidation state of Hadean magmas and implications for early Earth’s atmosphere**

- Dustin Trail, E. Bruce Watson & Nicholas D. Tailby

See also

- News & Views by Scaillet & Gaillard

- **Subduction dynamics and the origin of Andean orogeny and the Bolivian orocline**

- F. A. Capitano, C. Faccenna, S. Zlotnik & D. R. Stegman

- **Correlation between deep fluids, tremor and creep along the central San Andreas fault**

- Michael Becken, Oliver Ritter, Paul A. Bedrosian & Ute Weckmann

- **Energetics and the evolution of human brain size**

- Ana Navarrete, Carel P. van Schaik & Karin Isler

See also

- [News & Views by Potts](#)

- **A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma**

- Corine Bertolotto, Fabienne Lesueur, Sandy Giuliano, Thomas Strub, Mahaut de Lichy

- [+ et al](#)

- **A novel recurrent mutation in MITF predisposes to familial and sporadic melanoma**

- Satoru Yokoyama, Susan L. Woods, Glen M. Boyle, Lauren G. Aoude, Stuart MacGregor

- [+ et al](#)

- **Alternatively activated macrophages produce catecholamines to sustain adaptive thermogenesis**

- Khoa D. Nguyen, Yifu Qiu, Xiaojin Cui, Y. P. Sharon Goh, Julia Mwangi [+ et al](#)

See also

- [News & Views by Whittle & Vidal-Puig](#)

- **Lyn is a redox sensor that mediates leukocyte wound attraction *in vivo***

- Sa Kan Yoo, Taylor W. Starnes, Qing Deng & Anna Huttenlocher

- **Image-based genome-wide siRNA screen identifies selective autophagy factors**

- Anthony Orvedahl, Rhea Sumpter Jr, Guanghua Xiao, Aylwin Ng, Zhongju Zou

- [+ et al](#)

- **Nuclear PKM2 regulates β -catenin transactivation upon EGFR activation**

- Weiwei Yang, Yan Xia, Haitao Ji, Yanhua Zheng, Ji Liang

- [+ et al](#)

- **Control of *Drosophila* endocycles by E2F and CRL4^{CDT2}**

- Norman Zielke, Kerry J. Kim, Vuong Tran, Shusaku T. Shibutani, Maria-Jose Bravo

- [+ et al](#)

- **Cascades of multisite phosphorylation control Sic1 destruction at the onset of S phase**

- Mardo Kõivomägi, Ervin Valk, Rainis Venta, Anna Iofik, Martin Lepiku

- [+ et al](#)

CORRIGENDUM

Top

- **DNA demethylation for hormone-induced transcriptional derepression**

- Mi-sun Kim, Takeshi Kondo, Ichiro Takada, Min-young Youn, Yoko Yamamoto [+ et al](#)

THIS WEEK

EDITORIALS

CLIMATEGATE2 More leaked e-mails from East Anglia, much less interest **p.6**

WORLD VIEW Psychologists must learn to share their data **p.7**



ENDOSYMBIOSIS Pioneer of cell biology Lynn Margulis dies **p.11**

Half way there

Spain, Italy and Greece all have new governments and new research laws. Despite the pressures of economic austerity, investing in science now could bring disproportionate benefits.

Three of the Mediterranean countries hit hardest by the debt crisis — Spain, Italy and Greece — have little in common beyond a shared coastline and an inappropriately low level of investment in research and development. But in the past few years, all three have shown a desire to reform their sclerotic research systems. And in the past month, each has acquired, one way or another, a new government.

The priority for these governments is to haul their countries back from the brink of disaster, and thereby help to prevent the collapse of the euro. Given the enormity and international significance of these endeavours, does it make any sense to lobby for science to be favoured, financially and politically, in the tough austerity packages the new governments will have to enact?

It does, for two reasons. First, any developed country without a reasonable science base faces a bleak future — a familiar mantra, but true. Scientists in all three countries have for years seen little national money available for research projects, and almost no new academic recruitment. The best scientists have just about survived on international grants, particularly from the European Commission. Greek and Spanish researchers have the most to complain about, as austerity measures have cut into their pay packets, along with those of other civil servants in their countries. It is not hard to imagine that further cuts could lead to a dangerous level of demoralization.

Second, all three countries are now somewhere along the process of enacting or implementing new laws governing how research is organized and evaluated, which will bring them into line with scientific norms elsewhere in Europe. There is every reason to devote political resources to ensuring that these reforms are carried through properly — and, given that it won't cost much, little reason not to.

These Mediterranean countries have tended to be opaque in their science funding and academic recruitment, and so cronyism has often been able to rule over meritocracy. The new laws should help to fix this, primarily by introducing peer review and evaluation. In Spain and Greece, the new laws would also introduce for the first time much-needed independent national agencies for the competitively allocated funding of basic research, along the lines of the European Research Council or the US National Science Foundation.

These laws have taken years of discussion — Greece's long-promised science law has not even been approved yet, although its university law was passed in August. In Italy, a law for research was approved at the end of 2009 and a law for universities one year later. Spain's science law was approved in June.

All these laws differ in scope and detail, and they are not perfect in every clause. In all three countries, for example, most academics will remain civil servants with jobs for life, disappointing those who had hoped that universities and publicly funded laboratories would gain more hiring flexibility. But whatever their shortcomings, if appropriately implemented, each of the laws will make science higher quality and better value for money.

All this matters even more because a handful of internationally competitive institutes, along with smaller pockets of excellence, have sprung up in each of these three countries, despite a lack of political support. The leaders of these institutes have chosen to operate through meritocracy; imagine how much will be achieved when this approach becomes mandatory for the entire scientific enterprise in these countries.

"Pockets of excellence have sprung up in each country, despite a lack of support."

Improvements in science in southern Europe will not only benefit the individual countries in which they occur, but will make Europe as a whole more competitive. Yet without new money, the legal frameworks may not be able to work the wonders expected of them. A new research funding

agency won't be much help without a budget.

Now is not the time to expect huge increases in science investment, but small increases could make a disproportionate difference. Half way to reform, science in Spain, Italy and Greece needs to be supported. Like fiscal reform, it promises a long-term pay-off for those countries, and for the continent. ■

False economy

The Danish government's plan to axe technology assessment is ill-conceived.

Liberal democracy and science combined so successfully in the twentieth century that the nations and societies in which they were strongest rose to economic and cultural dominance. Nothing suggests that the recipe might lose its appeal as the twenty-first century proceeds, with democracy and science gaining ground in parts of the Arab world and some developing countries. But advances in science also raise ethical and environmental concerns that need to be taken seriously.

Denmark, a small but technologically advanced country that in January will take over the rotating presidency of the European Union, has pioneered the use of participatory methods to assess the risks and societal impact of new technologies. But plans to disband the Danish Board of Technology (DBT), which has been a leader in studying public views and expectations of science and technology, suggest that Denmark's new government is not quite aware of the country's formidable tradition in democratizing science. If it proceeds with the plan — which, ironically, is intended to preserve funds for research — it risks destroying a critical piece of Europe's science-policy system.

The DBT was set up in 1986 to advise the Folketinget, the Danish parliament, but its work is not confined to Denmark. In 2009, for example, it carried out the most extensive study to date of how communication affects global public attitudes towards climate change, covering 4,000 citizens in 38 countries. Unlike simple opinion polls, the World Wide Views on Global Warming project gave participants extensive information on the science and economics of climate change (see www.wwviews.org). As the meetings progressed, widespread scepticism and doubt gave way to a high degree of consensus that climate change is real and should be dealt with promptly.

The DBT is currently assessing a range of other topics, including the sustainability of the Danish transport system, the risks related to synthetic biology, and the security of delivering government services using new information and communication technologies. On the international level, it is set to produce a policy report, World Wide Views on Biodiversity, for next year's Rio+20 summit on biodiversity in India.

That project and most ongoing assessments will perish if the board is dismantled. The governing coalition of three centre-left parties formed after the parliamentary elections in September intends to redirect the board's annual funding of about 10 million kroner (US\$1.8 million) to the 2012 research and education budget, which it is determined to spare from savage cuts. In times of financial crisis, such a stance deserves applause — but the impact on Danish society and government of the loss of competence in technology assessment and public engagement in science would far outweigh the benefits

to students and researchers of the modest sum gleaned from closing the DBT.

Many other countries take inspiration from Denmark on how to study and shape public attitudes towards the science that could help society address issues such as ageing, climate, energy and biodiversity. The DBT is leading a €5.4-million (US\$7.2-million) European

“The Danish government risks destroying a critical piece of the science-policy system.”

Union-funded project, called PACITA, to expand technology assessment based on public engagement (see www.pacitaproject.eu). Denmark's approach is also gaining a following in Asia, where demand for sustainable energy technology is growing rapidly. Policy-makers in China and South Korea, for example, are increasingly open to the participatory methods for technology assessment pioneered by Denmark.

Clearly, then, the DBT should be saved. A way forward might be for it to apply its analytical services more widely to international assessments, such as the biodiversity exercise it is currently engaged in. Domestically, it should serve all decision-makers, from municipal to national levels, rather than just parliament. Taking on more externally funded projects could help to compensate for unavoidable budget cuts.

Meanwhile, scientists and science academies in Denmark and abroad should voice their support for a democratic institution that has served science better than its small size and modest title might lead one to believe. Danish MPs and the government's young science minister, Morten Østergaard, should take note. ■

A poor sequel

Muted media response to the release of more climate e-mails shows science's strength.

In an interview with *Nature* about 12 months ago, to mark the first anniversary of the release of hacked e-mails in an incident now widely referred to as Climategate, climate scientist Phil Jones said he feared that the anonymous hackers were sitting on more material, and that they would release it. He also said, having been through the experience already, that if there were to be a repeat then he was confident he would deal with it much better second time around. Last week he was proved correct, on both counts.

In marked contrast to the original 2009 release, Jones and his employer, the University of East Anglia in Norwich, UK, this time responded rapidly and with a keen sense of what the media were going to be interested in. This was never a story about the integrity of climate science, but rather about the behaviour of those scientists whose e-mails painted an incomplete but troubling self-portrait. The absence of the chief protagonist last time around only fuelled the flames.

Within 24 hours of the second batch of e-mails being handed to climate-sceptic websites, Jones was in central London answering questions at a press conference. And the university made widely available its explanations of some of the excerpted messages doing the rounds on the blogosphere. The reasons for this change in approach range from the practical — the original Climategate broke on a Friday, whereas the sequel arrived on a Tuesday — to the personal — Jones was shaken badly in 2009 and in no state to face the press. There is also a generous dollop of hindsight. For almost two years, those involved have had to listen to criticism of how they handled Climategate and how they should have done it differently. Last week they were able to show that they have learned from their mistakes.

Still, the swift response from Jones and his university cannot alone explain the relative indifference with which the new batch of e-mails — Climategate 2 — has been met by the wider world.

Much of the media frenzy over the e-mail release in 2009 was prompted by the high political profile of climate change at the time. Nations were preparing for the heavily hyped and ill-fated Copenhagen summit, and soon afterwards a blunder in an Intergovernmental Panel on Climate Change report about melting Himalayan glaciers added to the sense of climate science not being what it seemed. Climategate 2 came just ahead of the latest round of United Nations climate talks in Durban, South Africa, but that meeting is lower profile. It is harder to draw attention when fewer eyes are on the subject.

There is also the sense that many in the media felt cheated by the original Climategate. They were led by the nose, by those with a clear agenda, to a sizzling scandal that steadily defused as the true facts and context were made clear. Many will not make the same mistake — to write headlines first and ask questions later — again. Plus, it is hard for anyone except the most committed conspiracy theorist to see much of interest in the content of the released e-mails, even taken out of context.

None of the independent investigations that followed the 2009 release found any hints of scientific misconduct. Critics won't find any in the new batch either — the animated discussions that the highlighted e-mails do include, not shy of strong personal opinions and the occasionally harsh judgement concerning the quality of this or that piece of work, never really stray from sound normal science.

Climategate did no lasting damage to science. In fact, it can be argued that the incident fostered climatology and improved the way the field is perceived by the general public. The anonymous onslaught — illegal and grossly low, to be sure — has forced scientists to speak more openly about the gaps, difficulties and uncertainties that they are facing. On a more general note, it served to remind scientists, and hopefully legitimate critics as well, that respect, honesty and politeness are essential qualities in any intellectual endeavour.

If anything, Climategate 2 may damage the cause of the climate sceptics who eagerly promote it. Despite their obvious lack of anything approaching credible evidence, their hyperbole, accusations, claims and allegations remain the same. Beyond the echo chamber they inhabit, who is still listening? You cannot, as Abraham Lincoln said, fool all of the people all of the time. And it is getting harder to fool them some of the time too. ■

➔ **NATURE.COM**
To comment online,
click on Editorials at:
go.nature.com/xhunqy



Psychology must learn a lesson from fraud case

Sharing data could help to avert scandals like the Diederik Stapel revelations, and improve the quality of research, says Jelte M. Wicherts.

In 1977, two jumbo jets collided at an airport on the island of Tenerife. The highly regarded pilot of a KLM plane was under pressure to depart, and ignored several signs and warnings that a Pan Am jumbo had not yet cleared the runway. He took off in mist without proper clearance and 583 people were killed. The disaster happened because of a failure of the checks and balances put in place to deal with low visibility, arrogance, stress and the tendency to neglect conflicting information. These measures have since been improved, and commercial aviation is safer today not because it pointed the accusing finger at the erring pilot, but because it drew hard lessons.

Psychology, my field of science, recently uncovered a disastrous case of prolonged misconduct by Diederik Stapel, a highly regarded scientist at Tilburg University in the Netherlands (see *Nature* 479, 15; 2011). In response, we too must critically consider the circumstances under which the misconduct took place, with the aim of improving checks and balances to avoid a repeat.

The committee that investigated Stapel's misconduct has yet to identify the research papers tainted by his fraud, but it has already noted that the closed culture that characterizes much psychology research greatly aided Stapel's deceptions. One may argue that his misconduct is exceptional, regardless of research culture. However, the minor transgressions that all scientists are tempted to commit, as pointed out by Jennifer Crocker in a World View last month (see *Nature* 479, 151; 2011), are more likely when there is less scrutiny.

The interim report of the investigating committee revealed that Stapel often refused to share his research data with colleagues, even co-authors on papers. To scientists in other fields, this may seem extraordinary; to psychologists it is sadly common practice.

In a 2006 study published in *American Psychologist*, I helped to show that almost three-quarters of researchers who had published a paper in a high-impact psychology journal had not shared their data (J. M. Wicherts *et al.* *Am. Psychol.* 61, 726–728; 2006). Several data sets, authors said, had been misplaced, whereas others were kept secret because they were part of ongoing work, or because of ethical rules meant to protect participants' privacy. Such confidentiality has long been the most common excuse that psychologists offer for not sharing data, but in practice, most simply fail to document their data in a way that allows others to quickly and easily check their work. It is not unusual for data that are shared to list variables only as VAR00001 through VAR00019, with no further explanation.

It is not just misconduct that flourishes in such secrecy. So too do the common and more insidious failings of error and bias in data analysis — for

example, the use of incorrect tests, reporting errors due to similarly named variables, favouring results that confirm a hypothesis and overly positive reporting of statistical outcomes.

It is striking that psychology researchers go to great lengths to blind their data collection to possible confounding effects, including the expectations of participants, observers and experimenters, but seem oblivious to problems in the subsequent analysis and reporting. In aviation, a co-pilot checks the pilot's every step, and actions are recorded in a 'black box' to reconstruct any errors. In psychology, co-authors rarely verify a study's analysis, which is effectively conducted inside a box. Readers of published papers are shown dense summaries of results — without proper data archiving, they can only hope for the best.

Psychology's culture of secrecy produces substandard science.

Reanalyses of statistics in published psychology papers show frequent errors, and the more reluctant authors are to share their data, the more likely it is that their papers will contain mistakes. Or to put it another way — the results that most need checking cannot be checked.

How to lift the veil of secrecy? Obligatory archiving of raw data in online appendices to journal articles or in repositories should be a precondition for publication. This would not only help to uncover misconduct — curious patterns in Stapel's data led to his downfall — but would also help to prevent and later correct honest mistakes and unduly positive reporting.

With online publishing, data can often be published alongside the researchers' chosen statistical analyses and their summary of the results. As part of the growing concern over scientific openness,

grant-giving organizations, academic publishers and professional organizations including the American Psychological Organization are already considering such options. Yes, there are practical problems, including the need to keep data or participants confidential, but these can be solved by embargoes on releasing data for longitudinal studies, guidelines for preprocessing raw data, proper anonymity and exemptions where necessary.

As a first step towards this new way of publishing results, my close colleagues and I have implemented a 'co-pilot' model for our statistical analyses, in which we share data between us for double-checking and preventing embarrassing errors. And perhaps the wider mood is changing. Last week, a colleague, who had previously been reluctant to share, sent me the data from his latest work without me asking. And the analyses proved excellent. ■

Jelte M. Wicherts is in the Department of Psychology, University of Amsterdam, Weesperplein 4, 1018 XA Amsterdam, the Netherlands. e-mail: J.M.Wicherts@uva.nl

TO SCIENTISTS
IN OTHER FIELDS,
NOT SHARING
DATA MAY SEEM
EXTRAORDINARY;
TO PSYCHOLOGISTS
IT IS SADLY
COMMON PRACTICE.

► **NATURE.COM**
Discuss this article
online at:
go.nature.com/9gcjrg

Stretchable Graphene Transistors with Printed Dielectrics and Gate Electrodes

Seoung-Ki Lee,^{†,‡,∇} Beom Joon Kim,^{§,∇} Houk Jang,^{†,‡} Sung Cheol Yoon,^{||} Changjin Lee,^{||} Byung Hee Hong,[⊥] John A. Rogers,[#] Jeong Ho Cho,^{*,§} and Jong-Hyun Ahn^{*,†,‡}

[†]SKKU Advanced Institute of Nanotechnology (SAINT) and Center for Human Interface Nano Technology (HINT), Sungkyunkwan University, Suwon 440-746, Korea

[‡]School of Advanced Materials Science and Engineering, Sungkyunkwan University, Suwon 440-746, Korea

[§]Department of Organic Materials and Fiber Engineering, Soongsil University, Seoul 156-743, Korea

^{||}Advanced Materials Division, Korea Research Institute of Chemical Technology, Daejeon 305-600, Korea

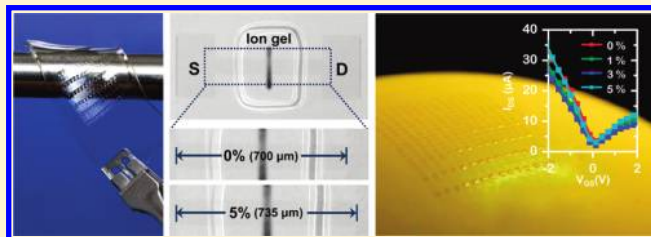
[⊥]Department of Chemistry, Seoul National University, Seoul 151-747, Korea

[#]Department of Materials Science and Engineering, University of Illinois at Urbana–Champaign, Urbana, Illinois 61802, United States

S Supporting Information

ABSTRACT: With the emergence of human interface technology, the development of new applications based on stretchable electronics such as conformal biosensors and rollable displays are required. However, the difficulty in developing semiconducting materials with high stretchability required for such applications has restricted the range of applications of stretchable electronics. Here, we present stretchable, printable, and transparent transistors composed of monolithically patterned graphene films. This material offers excellent mechanical, electrical, and optical properties, capable of use as semiconducting channels as well as the source/drain electrodes. Such monolithic graphene transistors show hole and electron mobilities of 1188 ± 136 and 422 ± 52 $\text{cm}^2/(\text{Vs})$, respectively, with stable operation at stretching up to 5% even after 1000 or more cycles.

KEYWORDS: Graphene transistor, stretchable devices, ion gel gate dielectric, printing process, low-voltage operation



Stretchable and transparent electronics can enable innovative classes of electronic applications for human interface technology.^{1,2} Such applications require electronic materials that can simultaneously exhibit excellent mechanical robustness, electronic functionality, and optical transmittance under a high strain.^{3–5} A variety of semiconductor materials have been explored for accomplishing this goal, ranging from small molecules and polymers to inorganic materials with various structural forms such as wires, ribbons, and platelets.^{6,7} One of the most significant hurdles to achieving transparent and stretchable devices using conventional semiconducting materials arise from their intrinsic limitations like poor mechanical stretchability and optical properties. As an alternative, graphene offers outstanding electrical, mechanical, and optical properties, with quantitative characteristics that, in many cases, outperform those of other known semiconducting materials.^{8–13} Many of these properties offer unique benefit in high-performance conformal, stretchable electronic devices.^{14–18}

Although several recent studies report the fabrication of flexible field-effect transistors (FETs) on plastic substrate using graphene thin films,^{19–21} significant challenges remain in the fabrication of graphene field effect transistors (FETs) that meet mechanical specification demands for stretchable electronics.

The main difficulties are (i) to develop low temperature, printing processes for materials that form the channel region, gate insulator, and electrodes on soft substrates with high thermal expansion coefficients²² and (ii) to overcome intrinsic limitations of mechanical properties associated with conventional materials and circuits through development of new materials or device architectures.²³

In this Letter, we present a promising route to the fabrication of an all-graphene-based FET array on a stretchable rubber substrate using a low-temperature printing process. First, all device components, including channel region and S/D electrodes, were transferred onto a substrate and then the gate insulator and gate electrode were printed in a manner that avoids the need for high-temperature processes. Second, all-graphene-based transistors, the source(S)/drain(D) electrodes, and semiconducting channels of which were monolithically patterned from graphene film, were fabricated on rubber substrates without the use of traditional metal electrodes. The resulting monolithic devices display

Received: June 24, 2011

Revised: August 29, 2011

Published: October 05, 2011

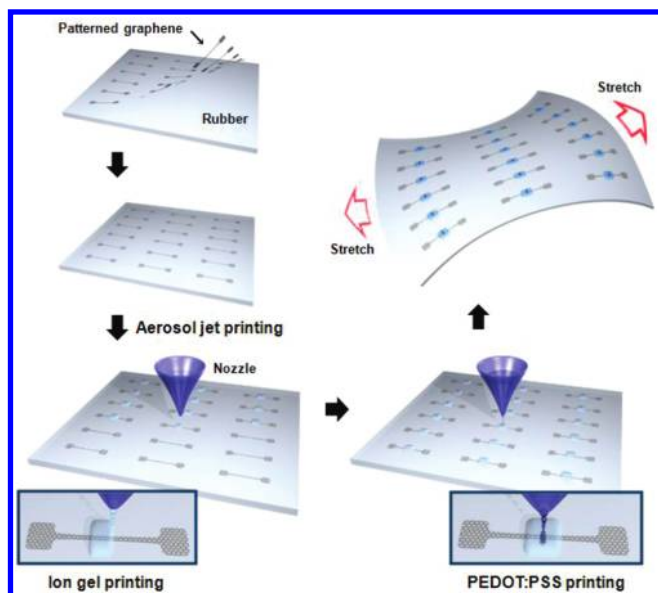


Figure 1. Schematic diagram of the fabrication procedure for arrays of graphene FETs on a stretchable substrate. Monolithically patterned graphene including source/drain electrodes and semiconducting channels were transferred on a rubber substrate with a supporting layer. After transfer, the supporting layer was removed with acetone. The ion gel gate dielectric and PEDOT:PSS gate electrodes were printed using commercial aerosol jet printing techniques.

several advantages, such as good mechanical stretchability, optical transmittance, and simple device design, as well as improved contact at the channel-to-S/D interface.

Figure 1 shows a schematic illustration of the steps involved in the fabrication of a graphene FET array on a rubber substrate. In the first step, high-quality monolayer graphene was synthesized on a large area Cu foil, as described previously¹⁷ and then the resulting material was stacked in a layer by layer fashion using direct transfer techniques (Figure S1, Supporting Information).¹⁸ The graphene patterns, including S/D electrodes and a semiconducting channel region, were defined by photolithography and oxygen plasma etching of graphene layers on Cu foil. The wide S/D region and the narrow long channel region were designed to optimize the effectiveness of the gate by adjusting its resistance. A photoresist (AZ 1512; commercial positive-type PR) supporting layer was spin-coated onto the graphene film on the Cu foil. The Cu metal catalyst was electrochemically etched using an aqueous 0.1 M ammonium persulfate solution. After etching, patterned graphene was transferred on rubber substrate as usual.¹⁵ For the low-temperature formation of the dielectric layer and gate electrode, a commercial aerosol jet printing technique was used (Figure S2, Supporting Information). The advantages of this jet printing method lie mainly in the high resolution (50 μm) and compatibility with various functional inks.^{22,24} Ion gel ink consisting of poly(styrene-methyl methacrylate-styrene) (PS-PMMA-PS) triblock copolymer and 1-ethyl-3-methylimidazolium bis(trifluoromethylsulfonyl)imide ([EMIM][TFSI]) ionic liquid in ethyl acetate solvent (weight ratio of 0.1:0.9:9) was printed onto the patterned graphene. Upon solvent evaporation, PS-PMMA-PS triblock copolymer in the [EMIM][TFSI] ionic liquid formed a well-defined physical gel through noncovalent association of PS components that were insoluble in the ionic liquid. In the last step, a poly(3,4-ethylenedioxythiophene) oxidized with

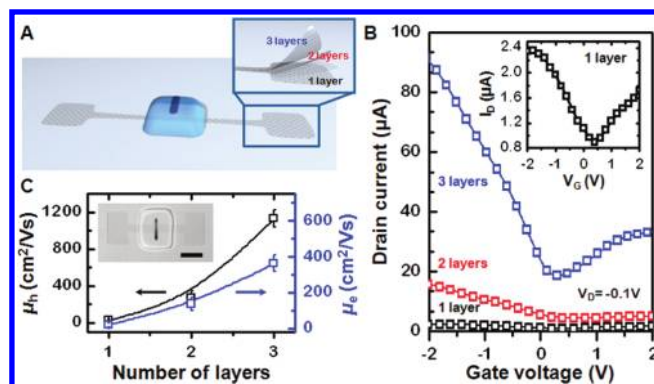


Figure 2. Schematic illustration and electrical characteristics of high-performance, transparent, stretchable graphene FETs on a PDMS rubber substrate. (A) Schematic diagrams of mono-, bi-, and tri-layer graphene FETs. (B) Transfer characteristics of graphene FETs as a function of the number of graphene layers. (C) Hole and electron mobilities of graphene FETs as a function of the number of graphene layers. The inset shows the microscopy image of tri-layer graphene FETs on PDMS (scale bar, 300 μm).

poly(4-styrenesulfonate) (PEDOT:PSS) gate electrode was printed over the channel. The conductivity of the printed PEDOT:PSS was measured to be 100 S/cm.

The outstanding mechanical properties of the graphene enabled device fabrication on rubber substrates, in ways that are not possible with conventional inorganic semiconducting materials. As a demonstration (Figure 2A), we fabricated high-performance, stretchable, and printable graphene FETs with high optical transmittance on poly(dimethylsiloxane) (PDMS) rubber (Figure S3:left, Supporting Information). Figure 2B shows the drain current (I_D) as a function of gate voltage (V_G) of the different layered graphene FETs. All the devices operate within a low voltage region (± 2 V) with a high on-current. The ultrahigh capacitance of the ion gel gate dielectric yielded this low-voltage, high-current operation. The specific capacitance of the ion gel was measured to be 5.17 $\mu\text{F}/\text{cm}^2$ at 10 Hz, much larger than typical values for 300 nm thick SiO_2 dielectrics (10.8 nF/cm^2). Clear ambipolar behavior was observed in the gate dependence of I_D , where a positive and negative V_G region represents electron and hole transport, respectively. The Dirac voltage which has the minimum conductance, for all the transfer curves was found to be at almost zero.²⁵

Unlike conventional transistors, the operation of ion gel gated transistors is based on the formation of a high capacitance electric double layer (EDL), at both the ion gel/semiconductor and ion gel/gate electrode interfaces under an electric field. For example, upon application of a negative gate bias, negative surface charges on the gate electrode attract positive [EMIM] ions and form an electric double layer at the ion gel/gate interface. At the same time, negative [TFSI] ions segregate to the ion gel/graphene interface. The two EDL are formed at the gate/ion gel and the ion gel/graphene interfaces with charge-neutral diffuse layer in between. Since the capacitance of this diffuse layer is negligible, the interfacial capacitance of ion gel can be then modeled as a serial combination of an electric double layer capacitance (C_{EDL}) and a quantum capacitance of the graphene (C_q).²⁶ As a consequence, the potential drop across the two capacitors is given by

$$|V_G - V_{G,\min}| = \frac{h\nu_F\sqrt{\pi n}}{e} + \frac{ne}{C_{\text{EDL}}}$$

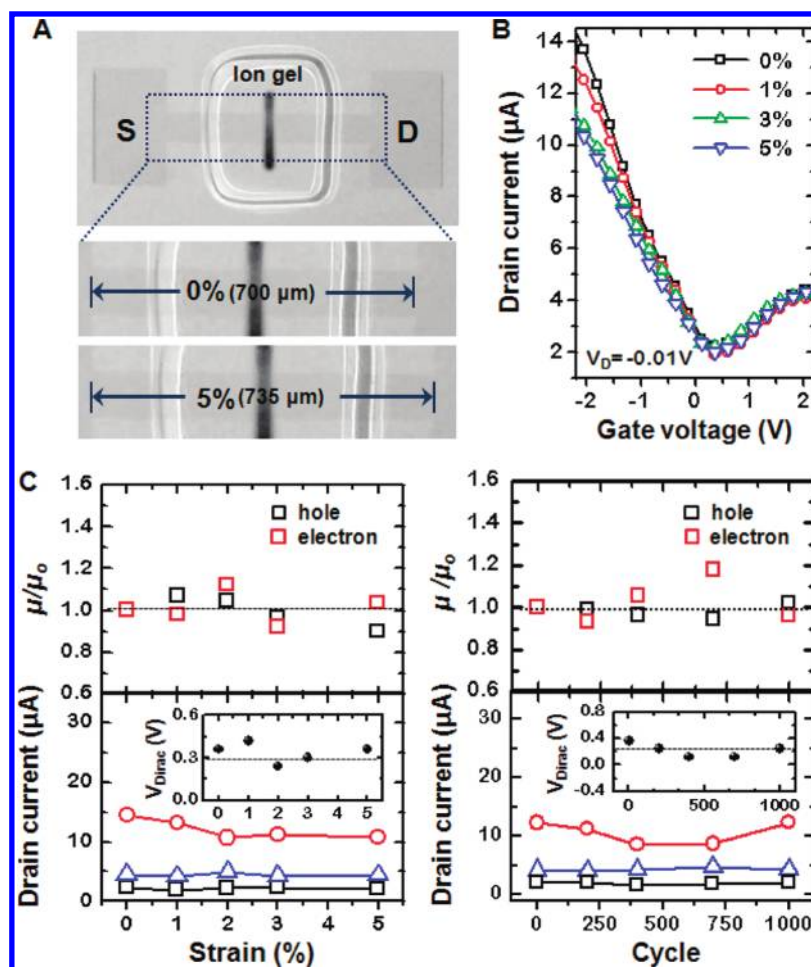


Figure 3. (A) Microscope images of tri-layer graphene FETs under substrate stretching up to 5% along the longitudinal direction of the channel. (B) Typical transfer characteristics of strained graphene FETs on PDMS. The device showed stable operation for stretching up to 5%, due to the excellent stretchability of the graphene and ion gel. (C) Normalized hole/electron mobilities, current levels (maximum hole (red circle)/electron (blue triangle), and the minimum current at the Dirac voltage (black square)) and Dirac voltage of graphene FETs as a function of the stretching level (left) and cycle during 3% device stretching and release (right).

where h is the reduced Planck's constant, v_F is Fermi velocity (1.1×10^6 m/s), e is the electron charge, and n is the charge density. On the basis of this equation, n was calculated by V_G , and the conductivity (σ) was plotted as function of n as shown in the right of Figure S3. From the slope of linear regimes in the σ vs n plot, the carrier mobility (μ) was calculated using $\mu = (d\sigma/dn)/e$. A quick formation of these two electric double layers is possible regardless of the area of overlap between the gate electrode and the active channel because it relies on the direct motion of highly mobile ions.²² As a consequence, unconventional device configurations in which the gate electrode does not fully cover the entire area of the graphene channel is possible (Figure 1).

First, monolayer graphene FETs were fabricated on PDMS substrate. The inset of Figure 2B shows the ambipolar transfer characteristics of the monolayer graphene FETs at $V_D = -0.1$ V. The average hole and electron mobilities were calculated as 26 and 20 $\text{cm}^2/(\text{V s})$, respectively. The poor carrier mobilities of the graphene FETs on PDMS substrate can be attributed to the both scattering from molecular species, such as moisture or other chemicals contained in the porous network structure of the PDMS²⁷ and exposing the graphene devices to contaminating species that can significantly reduce the transport properties.

To minimize the effects of PDMS-bound molecules, we fabricated bi- and tri-layer graphene-based FETs using multiple stacking method (Figure 2A). In these double- and tri-layer geometries, the bottom graphene layer provided sufficient screening of charged molecules on the substrates, acting as a barrier film; the top graphene layer was insensitive to the presence of moisture or other chemicals on the PDMS surface. In general, highly oriented multilayer graphene with an ABAB stacking sequence, as in graphite crystals, produces transistors with a lower on-current and on/off ratio due to interlayer coupling effects that become more prominent as the number of layers increases.²⁸ On the other hand, multilayer graphene films formed by transfer processes display weak interlayer coupling between layers due to the random orientations of the layer structures.¹⁷ As a result, device performance was dramatically improved compared with that achieved from monolayer graphene FETs (Figure 2B). Figure 2C shows hole and electron mobilities for different layered graphene FETs. The hole and electron mobilities increases with the number of graphene layers. The tri-layer graphene FETs yielded a hole mobility of $1131 \pm 96 \text{ cm}^2/(\text{V s})$ and an electron mobility of $362 \pm 45 \text{ cm}^2/(\text{V s})$, an order of magnitude greater than the corresponding values of the monolayer graphene FET.

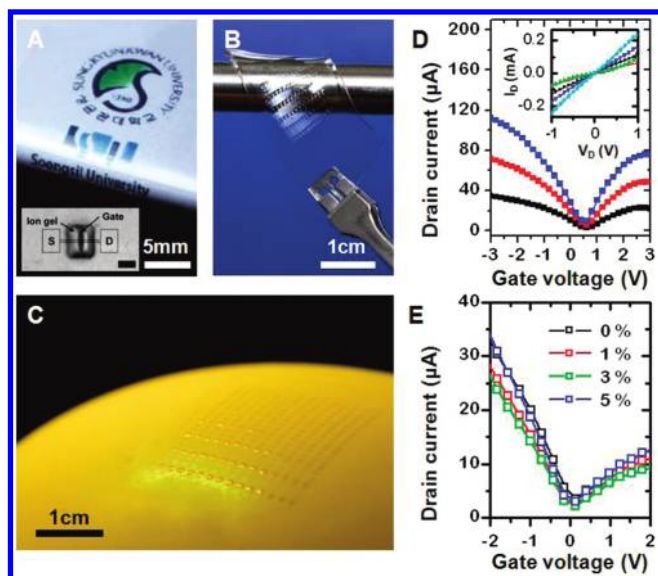


Figure 4. (A–C) Various images of the ion gel-gated graphene FETs on different substrate (polyethylene terephthalate (PET), PDMS and balloon). The inset shows the microscopy image of monolayer graphene FETs on PET (scale bar, 300 μm). (D) The transfer characteristics at three different V_D (–0.1 V (black), –0.2 V (red), –0.3 V (blue)) of monolayer graphene FETs fabricated on PET (channel widths of 20 μm and lengths of 500 μm). The hole and electron mobilities were 958 ± 237 and 512 ± 165 $\text{cm}^2/(\text{V s})$, respectively. Insert shows output characteristics of graphene FETs. In output curve, the gate voltage was varied between +2 V and –2 V in steps of –1 V. (E) Transfer characteristics of the tri-layer graphene FETs fabricated on a rubber balloon subjected to at most 5% strain ($V_D = -0.1$ V). The hole and electron mobilities were 435 ± 67 and 211 ± 38 $\text{cm}^2/(\text{V s})$, respectively, at 0% strain.

Uniaxial tunable tensile strain was applied to tri-layer graphene FETs with ion gel gate dielectrics by stretching the PDMS along the longitudinal direction of the channel. The device array was placed in a home-built uniaxial stretcher, with one side of the substrate fixed and the other side pulled to stretch the device (Figure S4, Supporting Information).

Figure 3A shows optical microscopy images of a graphene FETs at a device substrate stretching of 5%. FET performance was measured under 0–7% strain (Figure 3B and Figure S5, Supporting Information). The device exhibited stable operation at a strain of up to 5% due to the excellent stretchability of the graphene and ion gel. The hole and electron mobilities were calculated as 1188 ± 136 and 422 ± 52 $\text{cm}^2/(\text{V s})$, respectively, at 0% strain. The extensions of less than 5% yielded hole and electron mobility changes of at most 15%. Fatigue tests were performed on the graphene transistors. The electrical properties were invariant even after 1000 cycles of 3% stretching along the longitudinal direction (right frame of Figure 3C). The maximum hole and electron currents and the minimum current at the Dirac voltage were independent of stretching and fatigue (Figure 3C). The Dirac voltage did not shift significantly as shown in the inset of Figure 3C. However, degradation of device characteristics and broadening of the normalized mobility distribution were observed at large applied strains (above 5%), due at least partly to microcracks initiated by grain boundaries and/or by other defects in the graphene films (Figures S5 and S6, Supporting Information).²⁹

The good mechanical characteristics of graphene films enable integration onto highly compliant substrates, such as an ultrathin

plastic and a rubber material that are incompatible with conventional inorganic material-based electronics. As demonstrations, we fabricated graphene FETs on foldable plastic foil (Figure 4A, D), and freely deformable substrates such as rubber sheets and balloons (Figure 4B,C). Figure 4E shows the change in electrical properties of the graphene FETs during uniaxial stretching during inflation of the balloon. Graphene FETs on the balloon display excellent electrical properties without significant change during stretching. These results suggest that the graphene-based FETs can provide stable and robust operation during large-scale uniaxial and biaxial strain without any including special device designs, such as wavy or buckled configurations.

In summary, stretchable, transparent, monolithic graphene-based devices were realized by utilizing chemical vapor deposition grown graphene for both the semiconducting channel and the source/drain electrodes via low temperature printing procedures. This type of fabrication, without high temperature and vacuum processes, combined with intrinsically stretchable graphene films provides a promising route to creating electrical, optical, and mechanical performance for future stretchable electronic applications which would be difficult to achieve using conventional electronic materials. Although the stretchability in the reported designs are moderate, it might be possible to expand the range through optimized device structures.

■ ASSOCIATED CONTENT

S Supporting Information. Additional information on growth and transfer of graphene film and device fabrication. This material is available free of charge via the Internet at <http://pubs.acs.org>.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ahnj@skku.edu and jhcho94@ssu.ac.kr.

Author Contributions

[▽]S.-K. Lee and B. J. Kim contributed equally to this work.

■ ACKNOWLEDGMENT

This work was supported by the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2009-0082608, 2011-0003149, and 2011-0006268) and a Grant-in-Aid for Industrial Source Technology Development Programs from the Korea Ministry of Knowledge Economy (No. 10033309).

■ REFERENCES

- (1) Sekitani, T.; Noguchi, Y.; Hata, K.; Fukushima, T.; Aida, T.; Someya, T. *Science* **2008**, 321, 1468–1472.
- (2) Kim, D.-H.; Lu, N.; Ma, R.; Kim, Y.; Kim, R.; Wang, S.; Wu, J.; Won, S.; Tao, H.; Islan, A.; Yu, K.; Kim, T.; Chowdhury, R.; Ying, M.; Xu, L.; Li, M.; Chung, H.; Keum, H.; McCormick, M.; Liu, P.; Zhang, Y.; Omenetto, F.; Huang, Y.; Coleman, T.; Rogers, J. A. *Science* **2011**, 333, 838–843.
- (3) Lacour, S. P.; Jones, J.; Wagner, S.; Li, T.; Suo, Z. *Proc. IEEE* **2005**, 93, 1459–1467.
- (4) Rogers, J. A.; Someya, T.; Huang, Y. *Science* **2010**, 327, 1603–1607.
- (5) Park, K.; Lee, D.; Kim, B.; Jeon, H.; Lee, N.; Whang, D.; Lee, H.; Kim, Y.; Ahn, J.-H. *Adv. Funct. Mater.* **2010**, 20, 3577–3582.

- (6) Xiao, Lin.; Chen, Z.; Feng, C.; Liu, L.; Bai, Z.; Wang, Y.; Qian, Li.; Zhang, Y.; Li, Q.; Jiang, K.; Fan, S. *Nano Lett.* **2008**, *8*, 4539–4545.
- (7) Sekitani, T.; Someya, T. *Adv. Mater.* **2010**, *22*, 2228–2246.
- (8) Geim, A. K. *Science* **2009**, *324*, 1530–1534.
- (9) Lin, Y.; Dimitrakopoulos, C.; Jenkins, K.; Farmer, D.; Chiu, H.; Grill, A.; Avouris, Ph. *Science* **2010**, *327*, 662.
- (10) Schwierz, F. *Nat. Nanotechnol.* **2010**, *5*, 487–496.
- (11) Lee, C.; Wei, X.; Kysar, J. W.; Hone, J. *Science* **2008**, *321*, 385–388.
- (12) Bunch, J. S.; Zande, A. M.; Verbridge, S. S.; Frank, I. W.; Tanenbaum, D. M.; Parpia, J. M.; Craighead, H. G.; McEuen, P. L. *Science* **2007**, *315*, 490–493.
- (13) Nair, R. R.; Blake, P.; Grigorenko, A. N.; Novoselov, K. S.; Booth, T. J.; Stauber, T.; Pere, N. M. R.; Geim, A. K. *Science* **2008**, *320*, 1308.
- (14) Kim, K.; Zhao, Y.; Jang, H.; Lee, S.; Kim, J.; Kim, K.; Ahn, J.-H.; Kim, P.; Choi, J.; Hong, B. *Nature* **2009**, *457*, 706–710.
- (15) Li, X.; Cai, W.; An, J.; Kim, S.; Nah, J.; Yang, D.; Piner, R.; Velamakanni, A.; Jung, I.; Tutuc, E.; Banerjee, S. K.; Colombo, L.; Ruoff, R. *Science* **2009**, *324*, 1312–1314.
- (16) Lee, Y.; Bae, S.; Jang, H.; Jang, S.; Zhu, S.; Sim, S.; Song, Y.; Hong, B. H.; Ahn, J.-H. *Nano Lett.* **2010**, *10*, 490–493.
- (17) Bae, S.; Kim, H.; Lee, Y.; Xu, X.; Park, J.; Zheng, Y.; Balakrishnan, J.; Lei, T.; Kim, H.; Song, Y.; Kim, Y.; Kim, K.; Özyilmaz, B.; Ahn, J.-H.; Hong, B. H.; Iijima, S. *Nat. Nanotechnol.* **2010**, *5*, 574–578.
- (18) Wang, Y.; Tong, S.; Xu, X.; Özyilmaz, B.; Loh, K. *Adv. Mater.* **2011**, *23*, 1514–1518.
- (19) Chen, J. H.; Ishigami, M.; Jang, C.; Hines, D.; Fuhrer, M.; Williams, E. *Adv. Mater.* **2007**, *19*, 3623–3627.
- (20) Eda, G.; Fanchini, G.; Chhowalla, M. *Nat. Nanotechnol.* **2008**, *3*, 270–274.
- (21) Kim, B. J.; Jang, H.; Lee, S.-K.; Hong, B. H.; Ahn, J.-H.; Jeong, H. *Nano Lett.* **2010**, *10*, 3464–3466.
- (22) Cho, J. H.; Lee, J.; Xia, Y.; Kim, B.; He, Y.; Renn, M.; Lodge, T.; Frisbie, C. *Nat. Mater.* **2008**, *7*, 900–906.
- (23) Kim, D. -H.; Ahn, J.-H.; Choi, W.; Kim, H.; Kim, T.; Song, J.; Huang, Y.; Liu, Z.; Lu, C.; Rogers, J. A. *Science* **2008**, *320*, 507–5011.
- (24) Lee, J.; Panzer, M.; He, Y.; Lodge, T.; Frisbie, C. *J. Am. Chem. Soc.* **2007**, *129*, 4532–4533.
- (25) Zhang, Y.; Tan, Y.-W.; Stormer, H. L.; Kim, P. *Nature* **2005**, *438*, 201–204.
- (26) Das, A.; Pisana, S.; Chakraborty, B.; Piscanec, S.; Saha, S.; Waghmare, U.; Novoselov, K.; Krishnamurthy, H.; Geim, A.; Ferrari, A.; Sood, A. *Nat. Nanotechnol.* **2008**, *3*, 210–215.
- (27) Randall, G.; Doyle, P. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 10813–10818.
- (28) Sui, Y.; Appenzeller, J. *Nano Lett.* **2009**, *9*, 2973–2977.
- (29) Grantab, G.; Shenoy, V. B.; Ruoff, R. *Science* **2010**, *330*, 946–948.

NEWS IN FOCUS

BUSINESS The pharmaceutical industry peers over the patent cliff **p.16**

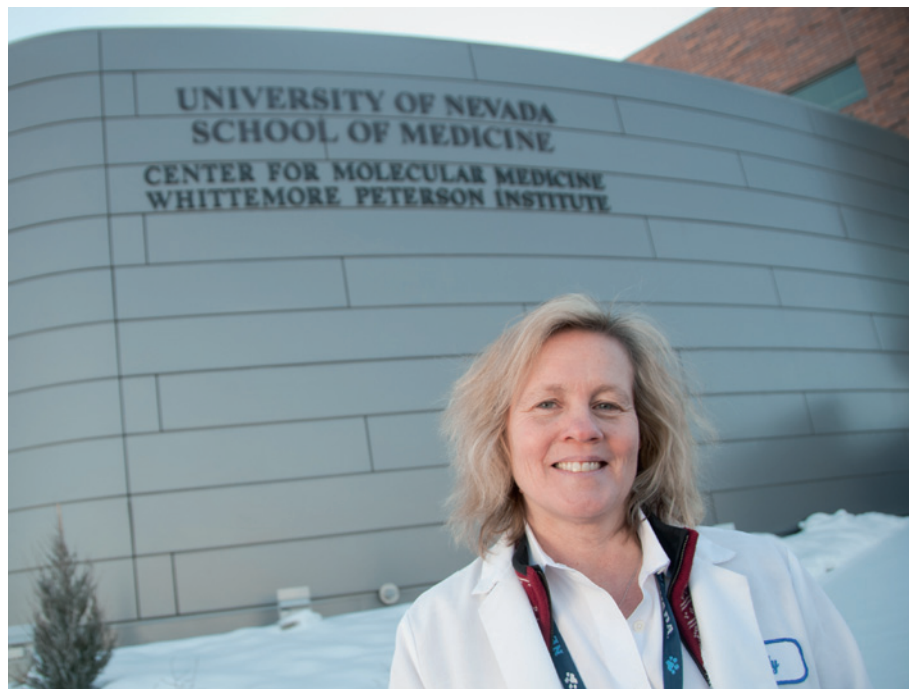
EUROPE Budget squeeze threatens fusion reactor and satellites **p.19**

ENERGY A greener way to drill in the Amazon rainforest **p.22**



BIOLOGY Inside the cell, an intricate landscape is revealed **p.26**

D. CALVERT/AP



Judy Mikovits has been accused of taking notebooks and other materials from her former employer.

RESEARCH

Embattled scientist in theft probe

Sacked virologist faces lawsuit from Nevada institute.

BY EWEN CALLAWAY

Controversy has dogged Judy Mikovits ever since she and her colleagues published evidence in *Science*¹ in 2009 suggesting that the retrovirus XMRV was linked to chronic fatigue syndrome (CFS). Yet as others dismantled the paper's claims, Mikovits has refused to recant her hypothesis, even after the paper was partially retracted².

Now the saga has taken a bizarre and shocking turn. On 22 November, Mikovits posted \$100,000 bail after spending four nights in jail in Ventura, California, as a "fugitive", according to a county-court docket. She is accused

of possessing stolen lab notebooks, a computer and other material belonging to the Whittemore Peterson Institute for Neuro-Immune Disease (WPI), a private research centre in Reno, Nevada, where she was research director. Mikovits faces extradition to Nevada, while the WPI is seeking the materials' return in a separate civil suit.

But in a letter to WPI lawyers dated 4 November, Lois Hart, a legal consultant to Mikovits, said that Mikovits did not have the notebooks, computer or any WPI intellectual property. Hart says the letter was posted

to the Internet without her permission, and has declined to comment on any other aspect of the affair. Mikovits did not respond to *Nature*'s questions. Her lawyer in the civil suit, Dennis Neil Jones, declined to comment or release any information telling Mikovits's side of the story, which has not yet been considered in court.

Tensions between Mikovits and the WPI reached breaking point two months ago, when the institute's president and chief executive, Annette Whittemore, fired Mikovits for refusing to share cell samples with a colleague.

Their relationship could not have been more different when *Nature* visited the WPI in January, to report on Mikovits's defence of her XMRV claims³. In interviews with Mikovits and Whittemore, there was no indication that their close friendship would soon implode. The two women ate breakfast together most mornings, and were optimistic about their research programme into the origins of CFS, a collection of symptoms including exhaustion and muscle pain that has no known cause.

But no published study to date has been able to reliably find XMRV in either CFS patients or healthy people. Some scientists believe that the original results are due to contamination, and a *Science* paper published in May⁴ suggests that XMRV emerged as a laboratory contaminant in the 1990s. One of Mikovits's former collaborators retracted his contribution to the 2009 *Science* paper on 22 September over concerns of contamination, although Mikovits declined to retract the whole paper. The journal is now investigating indications that data in the paper were misrepresented.

On 29 September, Whittemore dismissed Mikovits for refusing to share cell samples with Vincent Lombardi, another once-close colleague. That day, an incensed Mikovits told Max Pfost, a scientist in her lab who detected the first traces of XMRV in CFS patients, that "she had had enough of WPI" and that "WPI would go down", according to a 16 November affidavit signed by Pfost and filed by the WPI's lawyers to a court in Washoe County, Nevada. Mikovits rented a car that evening and drove to southern California, where she and her husband have a home, the affidavit says. Whittemore and Lombardi both declined to comment on the case.

Shortly after her dismissal, Mikovits told *Nature* that she planned to continue her research at another institution, supported by a grant of roughly \$1.5 million from the US National Institute of Allergy and Infectious

NATURE.COM
For a longer profile of
Judy Mikovits see:
go.nature.com/ei49zy

► Diseases (NIAID) that she won while at the WPI. But such R01 grants are awarded to institutions, not individuals: when a principal investigator (PI) leaves, the institution is allowed to choose whether to keep the grant and name a new PI, or transfer the grant to the original holder's new institution. Both actions require approval from the funding agency. The NIAID would not comment on the grant, which runs to August 2014.

Pfost has declined to answer *Nature's* questions, but according to his affidavit Mikovits gave him the keys to her office at a bar on the evening of 29 September and instructed him to retrieve laboratory notebooks and biological samples. Early the next morning, Pfost took 12–20 notebooks from a locked desk in Mikovits's office. The WPI noticed the notebooks were missing on 30 September, according to a representative for the institute. Claiming they were the property of the institute, it reported them missing to campus police. It was 17 October before Mikovits picked up the notebooks from Pfost, before returning to California.

Another worker in Mikovits's lab told lawyers in a signed affidavit on 21 November that after being fired, Mikovits asked her to ship cell lines and blood samples from the laboratory to Frank Ruscetti, a collaborator and former mentor to Mikovits at the US National Cancer Institute in Frederick, Maryland. She did not send the samples. Pfost, too, refused a request to send additional lab items to Ruscetti. Ruscetti has declined to comment, and there is no indication that he was aware of these events.

In a civil suit filed on 4 November, the WPI is seeking return of the notebooks, as well as a computer and flash drives that Mikovits used to store lab data and patient records. The institute obtained a restraining order preventing Mikovits from destroying, altering or disseminating any of the information contained in the materials. Mikovits is due back in court in California on 19 December to face potential extradition to Nevada, where she could face charges that she possessed stolen property. According to a WPI spokesperson, some of the missing materials have now been recovered.

Patients with CFS, also known as myalgic encephalomyelitis (ME), feel "very let down and very confused", says physician Charles Shepherd, medical adviser to the ME Association, based in Gawcott, UK.

"I think those people who believe in XMRV are going to continue to believe," he adds. "But I think it's going to be very difficult for Mikovits to continue this research." ■

1. Lombardi, V. C. *et al. Science* **326**, 585–589 (2009).
2. Silverman, R. H. *et al. Science* **334**, 176 (2011).
3. Callaway, E. *Nature* **471**, 282–285 (2011).
4. Paprotka, T. *et al. Science* **333**, 97–101 (2011).



Species such as the chambered nautilus would benefit from restrictions on activities in the Coral Sea.

ENVIRONMENT

Australia's marine plans questioned

Ocean reserves offer inadequate protection, critics say.

BY DANIEL CRESSEY

Home to pristine reefs, rare sharks and vast numbers of exotic fish, the Coral Sea is a unique haven of biodiversity off the northeastern coast of Australia. If a proposal by the Australian government goes ahead, the region will also become the world's largest marine protected area, with restrictions or bans on fishing, mining and aquaculture.

The Coral Sea reserve would cover almost 990,000 square kilometres and stretch as far as 1,100 kilometres from the coast (see map). Unveiled by environment minister Tony Burke on 25 November, the proposal would

be the last piece in a jigsaw of proposed marine reserves that stretch around Australia's coast.

But the scheme is attracting criticism from scientists and conservation groups, who argue that the government hasn't gone far enough in protecting the Coral Sea, or in other marine reserves in the coastal network.

Hugh Possingham, director of the Australian Research Council (ARC) Centre of Excellence for Environmental Decisions at the University of Queensland, points out that little more than half (507,000 km²) of the Coral Sea reserve is proposed as a 'no take' area, in which all fishing would be banned. The world's largest existing marine reserve, established last year by the

J. FREUND/NATUREPL.COM

British government around the Chagos Islands in the Indian Ocean, spans 544,000 km² and is a no-take zone throughout. An alliance of campaigning conservation groups, including the WWF and the Pew Environment Group, argues that more of the Coral Sea should receive this level of protection.

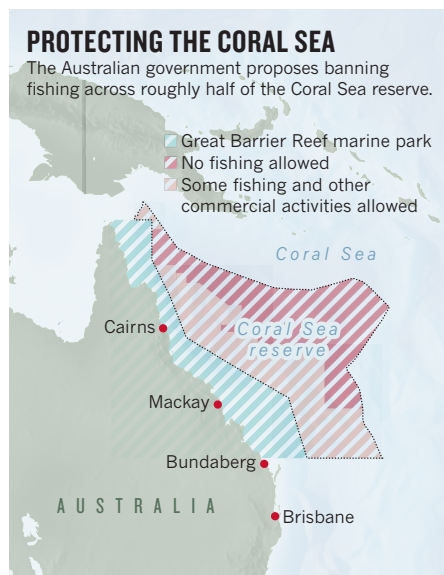
"I would like to have seen more protection for coral reefs," says Terry Hughes, director of the ARC Centre of Excellence for Coral Reef Studies at James Cook University in Townsville, Queensland. "More than 20 of them would be outside the no-take area and vulnerable to catch-and-release fishing".

As *Nature* went to press, the Australian government had not responded to specific criticisms of the plan. But Robin Beaman, a marine geologist at James Cook University, says that the reserve does "broadly protect the range of habitats" in the sea. Beaman's research helped to inform the government's proposal. "I can vouch for the huge effort that government agencies and other organizations have put into trying to understanding the ecological values of this vast area," he says.

Reserves proposed earlier this year for Australia's southwestern and northwestern coastal regions have also been criticized for failing to give habitats adequate protection. In August, 173 marine scientists signed an open letter to the government saying they were "greatly concerned" that the proposals for the southwestern region had not been based on the "core science principles" of reserves — the protected regions were not, for instance, representative of all the habitats in the region, they said.

Critics say that the southwestern reserve offers the greatest protection to the offshore areas where commercial opportunities are fewest and where there is little threat to the environment, a contention also levelled at the Coral Sea plan.

Public consultation on the proposal will conclude in February. ■



BUSINESS

Scientists, meet capitalists

US agencies and scientific societies aim to create jobs by teaching researchers how to be entrepreneurs.

BY EUGENIE SAMUEL REICH

It will be a presentation unlike any Satish Kandlikar has given before. At a meeting on 14 December, the mechanical engineer, who works at the Rochester Institute of Technology in New York state, will speak to venture capitalists at Stanford University in California about a technology to cool and extend the life of light-emitting diodes. He's not raising money, but — along with 20 other researchers — he'll be playing the role of entrepreneur in front of a roomful of instructors from the business world.

The moment is also a first for the Innovation Corps (I-Corps), launched in July by the US National Science Foundation (NSF). The programme has given Kandlikar and 20 other principal investigators US\$50,000 each to develop a business plan centred on their NSF-funded research. At the Stanford meeting, grant recipients will seek feedback on their ideas. Turning research into products is not a task his team has much experience with, says Kandlikar: "As researchers, we have no idea how to commercialize a product. The programme is the driving force."

US science agencies have long encouraged their grantees to build bridges with industry. Through the Small Business Innovation Research programme, founded in 1982, the NSF, the US Department of Energy and other agencies allocate 2.5% of their grant money each year to business ventures. Many US universities also help faculty members to patent and commercialize their research.

Now, with unemployment high and cuts to federal budgets looming, science agencies want to cast themselves as part of the economic solution, rather than expensive add-ons. Across government, agencies are finding ways to link their activities to job creation. The NSF's move to encourage academics to reinvent themselves as entrepreneurs, and perhaps found firms, takes the drive to commercialize to a new level. The I-Corps is already gearing up to receive its next round of applications in January.

The approach is spreading. Concerned

that US research jobs in chemistry are in decline, the American Chemical Society (ACS), based in Washington DC, has begun an entrepreneurship training programme, with a first round of applications due on 15 January. The idea is to try to get chemists to create their own opportunities.

"When my students say 'I want to get out and do something to help the world', I say basically you have to start your own company now," says George Whitesides, a prominent Harvard University chemist who chaired a recent ACS panel to study innovation, chemistry and jobs. Whitesides was speaking on 2 November at a meeting of the President's Council of Advisors on Science and Technology in Washington DC.

But many academics are reluctant to take a chance with a start-up, says Henry Sauermann, an economist at the Georgia Institute of Technology in Atlanta who has studied scientists' career choices. "Their main concern is lack of job security and stability," he says.

Sociologist Waverly Ding at the University of Maryland, College Park, who has studied academic entrepreneurship, adds that programmes such as the I-Corps may work well for established investigators. But younger scientists have a harder time raising capital, and taking time away from research is less likely to benefit them, she says.

Sauermann believes that entrepreneurship programmes should focus on educating academics about what such choices involve. Errol Arkilic, one of three NSF programme officers organizing the I-Corps, says this is the intention. That's why the venture capitalists on hand to assess business plans at the Stanford meeting later this month will be acting as instructors, not as potential investors.

"The programme was established to help researchers understand the commercial viability of their research," not to help them found companies and create jobs, says Arkilic — although he adds that the NSF would not be unhappy to see that as an outcome.

Kandlikar wants his work to yield economic benefits, but the effort is already taking its toll, with his group spending about 100 hours a week on its business plan. Like science, the market demands devotion. ■

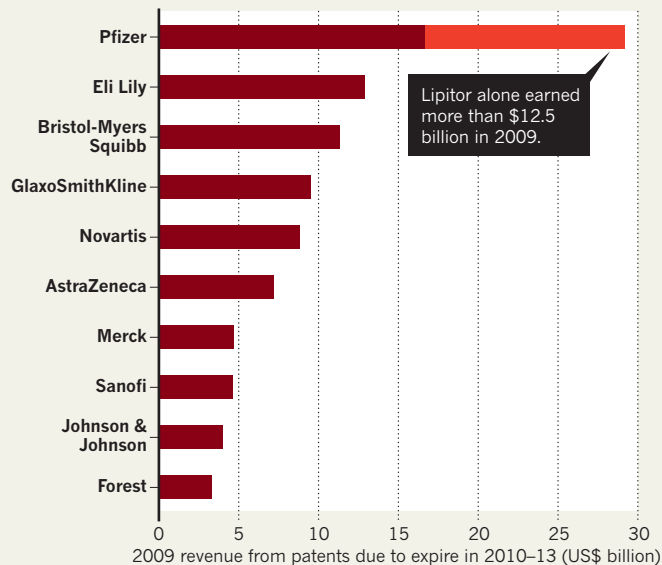
► **NATURE.COM**
For more on the
I-Corps see:
go.nature.com/mzyqmh

PEERING OVER THE PATENT CLIFF

Lipitor and other drugs with patents expiring between 2010 and 2013 generated more than US\$100 billion in sales in 2009 alone. Prospects are dim for offsetting this revenue with that from new drug discoveries in the near term.

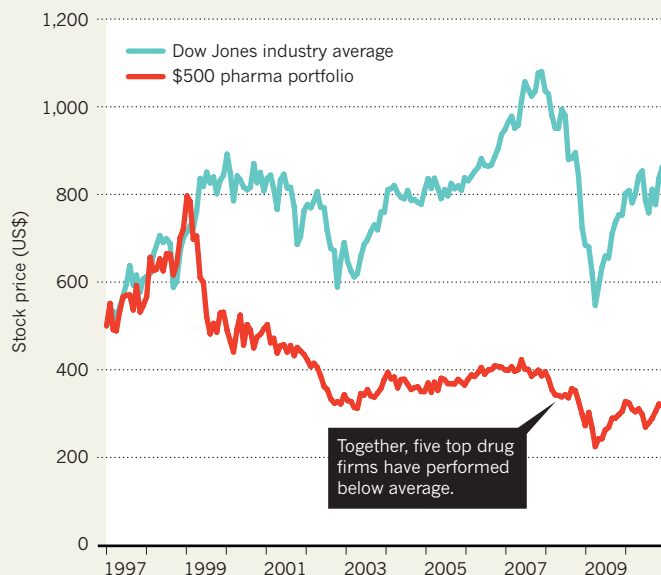
THE BIGGEST LOSERS

The top ten drug manufacturers facing leaner times thanks to patents expiring in 2010–13.



A DECLINING PORTFOLIO

A 1997 investment of \$500 equally divided between Pfizer, Eli Lilly, Bristol-Myers Squibb, GlaxoSmithKline and Novartis would have fallen by 39% by the end of 2010.



DRUG DISCOVERY

Blockbuster drug bows out

Pharmaceutical industry anxiously struggles to retool as Lipitor patent expires.

BY HEIDI LEDFORD

With sales of more than \$100 billion since it was introduced in 1997, the cholesterol-lowering drug Lipitor (atorvastatin) is an unparalleled pharmaceutical superstar. But as its patent expires on 30 November and its first generic competitor takes the stage, Lipitor is also a painful reminder of the challenge that such 'patent cliffs' pose for the big drug companies, including Lipitor's developer Pfizer, based in New York.

"The industry is truly in crisis mode," says Ken Kaitin, director of the Tufts Center for

the Study of Drug Development in Boston, Massachusetts. "Companies realize they have to change."

Including Lipitor, patent expiries in 2010–13 will jeopardize revenues amounting to more than US\$95 billion for ten of the largest drug companies (see 'Peering over the patent cliff'). But the real problem is much deeper, says Peter Tollman, managing director of the Boston Consulting Group. "The patent cliffs wouldn't be such a big issue if research and development productivity hadn't collapsed."

A decade ago, investors began to realize that the pharmaceutical industry, which once rewarded them with double-digit growth rates,

faced a perilous future. Someone buying \$100 worth of Pfizer stock in 1986 would have seen it grow in value to \$168 by 1999. But by the end of 2010, that investment would have shrunk to \$20. Research and development (R&D) budgets soared, but the rate at which new drugs were approved did not. Frederick Frank, vice-chairman of the Peter J. Solomon Company, an investment bank based in New York, has estimated that to sustain a 10% annual growth in sales, Pfizer would have to introduce a minimum of 11 new products, each earning about \$750 million, each year.

Investors recognized that the industry would not be able to sustain that kind of growth.



**MORE
ONLINE**

SPECIAL

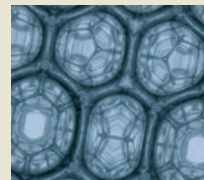


Climate showdown in Durban — *Nature* follows the efforts to shape a post-Kyoto world
go.nature.com/dw9pu2

MORE NEWS

- Computer software spots manipulation of digital images
go.nature.com/exizwx
- The problems with emissions trading
go.nature.com/iwnhvu
- Archaeologists land world's oldest fish hook
go.nature.com/e7teh1

FROM THE BLOG



Scientists make theoretical 'perfect' foam a reality
go.nature.com/3lsqbx

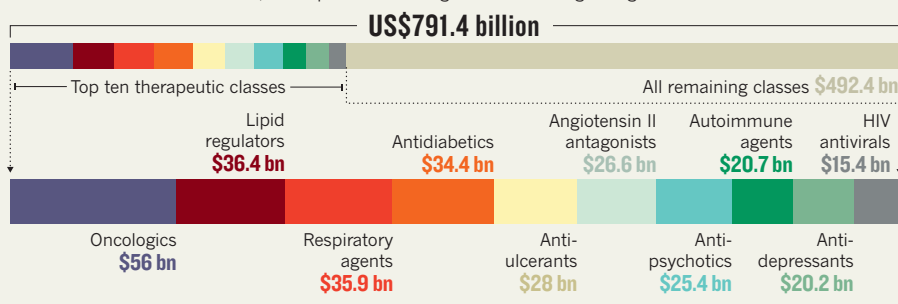
GROWTH LIMITS

The drug industry has more than doubled its investment in research and development since 1997, but the number of new products winning approval has failed to keep pace.



THE NEXT BLOCKBUSTER?

Blockbuster drugs typically emerge from therapeutic classes, such as cancer drugs, in which demand is high in wealthy countries (shown by 2010 revenue). Breakthroughs in classes with relatively few treatment options, such as for Alzheimer's disease, could produce a new generation of mega-drug.



SOURCE: PHRMA/FDA (TOP); IMS HEALTH

By 2000, pharmaceutical pipelines were filled with 'me too' drugs that offered low risk and a quick route to approval, but only marginal improvement over drugs already available. As regulatory standards rose and key markets became crowded with competitors, many of them cheap generics, the industry began to suffer from what some have called 'better than The Beatles' syndrome. "Imagine if you weren't allowed to publish music unless you were better than The Beatles," says Tollman. "And The Beatles were being given out for free while you had to sell yours at full price."

So shareholders walked away. "Investors looked at the industry and said, you're not doing anything innovative," says Kaitin. "At the end of the day, the odds of actually getting a product to market were getting worse, and investors started dropping out."

To cope with the worsening situation, drug companies have slashed R&D costs (see *Nature* 470, 154; 2011) and forged outside collaborations with other firms and with academia (see *Nature* 474, 433–434; 2011). Pfizer aims to trim its R&D budget by up to \$3 billion between 2010 and 2012, and the Michigan site that developed Lipitor was shuttered in 2007. Some drug companies, such as London-based GlaxoSmithKline, have boasted a renewed emphasis on risk-taking and innovation, to get away from the 'me too' era. Whether this strategy will pay off remains to be seen — it takes upwards of a decade to develop a new drug.

Kaitin notes that Eli Lilly, a relatively small company based in Indianapolis, Indiana, is taking the second-largest hit from patent

expiries in 2010–13. Not surprisingly, Lilly has dramatically revamped its R&D programme, forging new relationships with academia and contract-research organizations.

Could another round of mass-market blockbusters reverse the trend? Unlikely, say industry watchers. Companies are increasingly moving towards niche drugs based on personalized medicine, says Kaitin — although there are exceptions. Swiss drug-maker Roche, based in Basel, is developing a cholesterol drug called dalcetripib that has a different target from Lipitor, and which analysts predict could

"The patent cliffs wouldn't be such a big issue if research and development productivity hadn't collapsed."

bring in \$10 billion per year, notes Tollman. Companies are still highly focused on diabetes drugs with potentially large markets, and if an effective treatment is developed for Alzheimer's disease, "that's

going to do unbelievably well", he adds.

Even so, Kaitin cautions that new blockbusters are likely to be short lived. Lipitor owed its success in part to working better than other drugs, with few side effects, in a class where demand for effective therapies is high. Now, many companies are flocking to the same drug targets, which means that the first of a kind to be approved won't enjoy a monopoly for as long as older drugs did. "We won't see drugs with the longevity of Lipitor," Kaitin says. "We'll see drugs that sporadically reach very high levels of revenue, and then come down." ■



Europe's Earth-monitoring programme GMES is scheduled to begin launching a fleet of satellites in 2013, but its funding future is now uncertain.

POLICY

Outcry over EU budget plan

Costly Earth-monitoring and fusion-energy projects could be stripped from main budget.

BY DECLAN BUTLER

As Europe's financial crisis deepens, a storm is also brewing over proposals that would change how two giant science and technology projects are funded.

Both ITER — the international effort to build a fusion-energy test reactor — and an ambitious Earth-observation project called the Global Monitoring for Environment and Security (GMES) programme are too costly to remain under the general budget of the European Union (EU), according to proposals from the European Commission. The solution, it says, is to corral funding for both projects separately from the next general budget, which will span 2014–20.

The commission suggests that the projects — along with future large-scale science programmes — be supported through new intergovernmental organizations. EU member states would fund these bodies, perhaps along with an additional, capped contribution from the EU budget. This arrangement would reduce the main EU budget's exposure to the large cost overruns that are common in big science projects, the commission argues.

The plan has already sparked vigorous

opposition from several member states and from scientists. On 10 November, ministers from eight member states, including Germany, France and Italy, wrote to Máire Geoghegan-Quinn, the EU research commissioner, as well as the industry and budget commissioners, warning that the proposal “would only lead to the weakening of these two projects both in terms of funding and governance”. The proposal is likely to go through tortured negotiations between the commission, the Council of Ministers and the European Parliament right up until the 2014 budget settlement.

ITER is undoubtedly the highest-profile example of the strain that enormous projects can put on the EU budget. Crisis talks are scheduled for 1 December in Brussels to find a way to make up a €1.2-billion (US\$1.6-billion) hole in funding for ITER's construction. If a solution is not found in the coming weeks, work at the ITER site in Cadarache, France, could face major delays.

➔ NATURE.COM
For news of the
Horizon 2020
budget proposal see:
go.nature.com/Guegfo

In 2006, ITER was projected to cost €5 billion, but that sum has now grown to €15 billion, forcing the European Commission into a series

of financial contortions to make up repeated shortfalls. To insulate the rest of the EU budget from ITER's bloat, the commission is proposing that for 2014–20, Europe's anticipated €2.7-billion spending on ITER should be channelled through a separate fund.

The commission also wants to make a similar arrangement for GMES, which is expected to cost €5.8 billion over the same period. The programme, led by the EU with the European Space Agency (ESA) as a major partner, aims to launch a fleet of five families of ‘Sentinel’ satellites. Drawing on data from other existing or planned satellites and ground sensors, GMES would generate continuous, cross-calibrated, long-term data sets on the state of the planet and its atmosphere. The system would act as a weather service writ large, delivering images, maps, models and forecasts to users 24 hours a day, 365 days a year. Its capabilities would include monitoring and forecasting of climatic change, flood risks, soil and coastal erosion, crop and fish resources, air pollution, greenhouse gases, iceberg distribution and snow cover.

The commission is expected to release detailed proposals for GMES funding this week. But *Nature* has already learnt that the

plan would require the 27 EU member states to pay into a GMES fund, with contributions calculated pro rata according to each state's gross national income. Parallel proposals for a change to ITER's funding mechanism are expected on 13 December.

ON THE CUSP OF REALITY

The timing of the commission's proposals could hardly be worse for GMES, whose organizers plan to start launching its satellite fleet in 2013. "Europe, for once, has taken [world] leadership to develop a very comprehensive Earth-observation system, with multiple satellites and instruments at the technological cutting edge," says Josef Aschbacher, the head of GMES at ESA, based in Frascati, Italy. "We are now on the cusp of making this a reality."

He notes that the project has an exemplary record of staying on budget, even during its long and financially risky research and development phase. Once the satellites enter into operation in 2014, costs will become much more predictable and easier to control. But "we cannot launch the satellites if we have no guarantee that the operational funding and governance structure will be in place afterwards", Aschbacher says.

A new funding mechanism would require unanimous agreement among the EU member states and could take at least five years to

establish, he adds, posing a serious setback for the project or killing it altogether.

Aschbacher says that he is now "very concerned" for GMES's future, and notes that a working document accompanying the commission's budget proposals seemed to contradict the proposals, concluding that taking GMES out of the main EU budget entirely

"Europe, for once, has taken world leadership to develop a very comprehensive Earth-observation system."

before the programme was firmly established would "create high uncertainty for GMES" and "more than likely ... mean a discontinuation of GMES". Some observers speculate that the commission is using this potential threat to GMES as a bargaining chip to win concessions from member states on other issues, and that, ultimately, a compromise solution to its funding will be hammered out.

CREATIVE ACCOUNTING

Meanwhile, ITER's financial problems are much more pressing. As part of general budget negotiations in November, ITER was thrown an extra €100 million for next year, but efforts to find the remaining €1.2 billion to cover cost overruns have failed.

On 23 November, a 'trilogue' of the council, commission and parliament could not agree on a commission proposal to scrape the money together from other parts of the EU budget. It suggested transferring €300 million from the Framework research budget; €600 million from the agriculture budget; €240 million from the EU administration budget; and €25 million from the justice, home affairs and education budget.

The result was a stalemate, so another trilogue will be held on 1 December. One source close to the negotiations says he is "cautiously optimistic" that a solution will be found at this meeting. If not, says Michel Claessens, a spokesman for ITER, there will be "significant delays" to the construction of "nearly all the buildings of the ITER site". ■

CLARIFICATION

The News story 'Summit urged to clean up farming' (*Nature* **479**, 279; 2011) relied on a draft report supplied by the Commission on Sustainable Agriculture and Climate Change that stated that agriculture is the "single largest contributor to greenhouse-gas pollution". The final version of the report states that agriculture is "a significant contributor to greenhouse gas emissions".



THE ROADLESS WARRIOR

To save the Amazon, Bruce Babbitt wants to isolate islands of oil and gas production amid a sea of trees.

Bruce Babbitt ambles down a walkway flanked by manicured lawns, gleaming office trailers and tidy rows of housing. The drone of vehicles competes with Latin music playing softly in the distance. Only the occasional birdsong or glimpse of macaws serves to remind that this encampment is smack in the middle of the Peruvian Amazon. Babbitt looks into the distance, where the separation towers of a natural-gas facility rise above the trees and shimmer with industrial splendour in the evening sun. “It’s an amazing sight,” he says.

The Malvinas natural-gas plant might seem the ultimate insult to a largely unspoiled tropical paradise, particularly for a lifelong conservationist such as Babbitt, who served as Secretary of the Interior — responsible for managing much of the United States’ federal land and natural resources — under US President Bill Clinton from 1993 until 2001. But where others see blight, Babbitt sees a vision of the future. He looks past the pipes and pollution and focuses instead on what makes this project stand out: seen from the sky, Malvinas is an island of industrial activity in a sea of trees. There are no roads into the site; everything that enters or leaves Malvinas, including gas, rubbish, food and people, does so by plane, boat or underground pipeline.

BY JEFF TOLLEFSON

The design is called an offshore-inland development, and Babbitt thinks it might be the western Amazon’s only hope. Historically, roads have paved the way for uncontrolled development throughout the region. Without them, there can be no associated logging, squatting or large-scale invasion of the forest, as has happened around other oil and gas operations in the Amazon. “There’s a huge rush all across this region, and the place is going to be destroyed unless this model is embedded in all future discovery and operations,” Babbitt says. “This is a message that curiously does not have a messenger.”

It does now, in Babbitt. In partnership with the non-profit Blue Moon Fund based in Charlottesville, Virginia, Babbitt has used his prominence to urge governments in the region to require that any oil and gas development in the Amazon basin follow the offshore-inland model. His efforts run counter to industry, which fears regulation, and to some environmentalists, who want to avoid selling out to oil and gas producers that have a poor track record in the Amazon. But Babbitt’s campaign is an extension of the work he did in the US government, where he forged innovative policies that balanced business and

J. TOLLEFSON

conservation interests on issues ranging from endangered species to energy development in Alaska.

Now 73 years old, Babbitt is dedicating himself to protecting the Amazon. Working with Enrique Ortiz, a programme officer with Blue Moon, Babbitt has been shuttling back and forth between Peru and his home in Washington DC to promote the offshore–inland model and a broader goal of smart, controlled development in the Amazon.

The heart of their campaign is Malvinas, which began processing gas from the surrounding Camisea natural-gas field (see ‘Islands in the jungle’) in southeast Peru in 2004. Apart from one field in Brazil, Camisea is the only example of offshore–inland development in the Amazon. The roadless plan was implemented under pressure from environmental groups and the Inter-American Development Bank (IDB), but the idea received scant attention, in part because Camisea has been beset by environmental problems such as pipeline leaks. Babbitt and Ortiz, however, see potential here. They are promoting the idea of roadless development at a crucial time, as energy companies look to expand their operations in the Amazon. Earlier this year, Babbitt and Blue Moon were instrumental in blocking a plan to build a new pipeline from the Camisea field that could have spurred widespread road construction and deforestation.

“I’m not a very emotional guy, but I’ll tell you I got emotional about this,” Babbitt says, after delivering his first public speech promoting his vision for roadless development at an energy conference in Lima in September. “It was important to make an issue of it, to demonstrate that it was possible to mobilize opposition.”

DAMAGE CONTROL

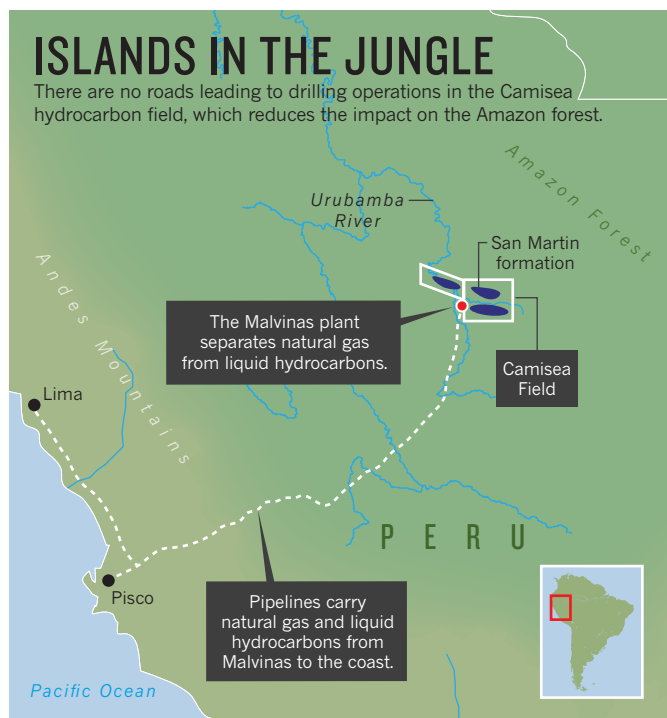
Malvinas’s isolation comes into focus as Babbitt’s helicopter rises into the sky and arcs out to the east, across the jungle. The Urubamba River borders the plant on the west, and a sea of green surrounds it on all other sides. Babbitt and Ortiz are accompanying a crew of six workers to San Martin Number 1, a platform for multiple wells tapping natural gas, propane, butane and other light hydrocarbons trapped in a sandstone formation thousands of metres below ground. En route, they pass over a clearing dotted with thatch-roofed huts near the Camisea River. It is home to one of several indigenous communities that have leased their land to a consortium led by the exploration and production firm Pluspetrol, based in Buenos Aires. ‘Uncontacted’ tribes that have yet to establish any ties with modern society roam the rolling hills on the horizon.

The helicopter circles the production site and gives Babbitt and Ortiz a good view of the operation. Apart from a few small structures containing production equipment, the main feature is an open grass field roughly the size of a football pitch, which is used for setting up equipment. Several pipes rise up out of the ground on one edge of the clearing then dive back below the surface at the base of a hill, marking the beginning of an underground journey to Malvinas, where gas will be separated from liquids then piped to the coast for distribution. Workers cleared a narrow path through the trees when they buried the pipeline, but the jungle has since reclaimed the land, completely camouflaging the route.

For the two environmentalists, Malvinas is a proof of principle. “This is far from an idyllic forest, but what I see is something that is contained,” Ortiz says. “As long as there is no oil or gas in the river, it’s just not that bad.”

Pluspetrol has several isolated production platforms tapping the eastern end of the Camisea field, and is expanding production with several platforms to the northwest. At each of these platforms, the company used horizontal-drilling techniques to create multiple wells that veer off underground and tap different regions, helping to limit the amount of forest that needed to be cleared.

The project’s remote location has made it difficult to assess what might have happened had roads been allowed into Camisea, but historically, deforestation has followed closely on the heels of road building¹. A lack of roads has kept the western Amazon relatively intact. To



the south, a classic fish-bone pattern of secondary roads can be seen branching off the InterOceanic Highway, and roads have accompanied oil and gas operations in Ecuador and northern Peru². One modelling study³ suggests that the southwestern Amazon, around the confluence of Peru, Brazil and Bolivia, could lose two-thirds of its forest cover by the middle of this century if roads continue to be built without additional protections.

CIRCLE OF LIFE

Babbitt started working full time on Amazon conservation issues just two years ago, after he retired from his law practice in Washington DC. He is not, however, new to the region. He made his first trip there in the summer of 1962, as a master’s student in geophysics at Britain’s Newcastle University. His professor, Keith Runcorn, pioneered the palaeomagnetic mapping that bolstered the case for plate tectonics, and Babbitt’s job was to collect rocks for palaeomagnetic analyses from transects up and down the eastern side of the Andes in Bolivia. He spent much of his time in camps run by Gulf Oil, a company based in Framingham, Massachusetts, that was busy doing early oil and gas reconnaissance.

Babbitt ended up abandoning rocks for law school and then politics, but his geological introduction to the eastern Amazon still comes in handy. “There’s a circular quality to life,” he says. “That first summer made an impression on me, and I made a commitment to get back.”

As a lawyer, Babbitt moved up through the ranks to attorney-general of Arizona, then to governor in the 1980s. He failed to win the Democratic nomination for the US presidency in 1988, but in 1993, Clinton appointed Babbitt to serve as secretary of the Department of the Interior, which put him in charge of roughly 200 million hectares of federal land. It was then that he began to think more deeply about low-impact ways to develop oil and gas, beginning with the North Slope in Alaska. In 1998, the interior department opened up around 1.6 million hectares of wilderness there for drilling but required the use of ‘ice roads’, which are created each winter to give heavy equipment access to the region. That meant that industrial activity was confined to the winter months and helped to prevent damage to the fragile tundra.

Babbitt started to focus on Camisea drilling in 2002, when he chaired a blue-ribbon commission established to review environmental policies at the IDB. At the time, a consortium was seeking US\$75 million in loans from the bank to build a pipeline from Camisea to the coast.



The Malvinas natural-gas facility has no roads leading into it.

Under pressure from environmentalists, the IDB imposed several requirements, including roadless development, collaboration with native communities and a programme to monitor biodiversity.

It was an important time for Babbitt. The Camisea example showed how international lending institutions could impose social and environmental constraints on such projects. And despite a litany of complaints about Pluspetrol's actions in Camisea, Babbitt and Ortiz say that the roadless requirements have held up — so far.

Their optimism is not shared by all. For many environmental and social activists, Camisea is a deeply flawed project. It has been rocked by scandal, from ruptured pipelines in the early years to questions about how the government distributes revenue from the project. Activist groups such as Amazon Watch in San Francisco, California, argue that Pluspetrol could have built fewer platforms by taking additional steps to consolidate their drilling, and that the pipeline has caused extensive erosion along the route carrying gas south out of the basin from Malvinas. César Gamboa, political director for the Law, Environment and Natural Resources centre in Lima, supports the offshore–inland model, and his group has received funding from Blue Moon to work on legal issues associated with oil and gas development and their impact on indigenous cultures. But Gamboa sees the model as just a first, small step and argues that companies need to do much more to limit the environmental and cultural impact of operations in the Amazon.

Gamboa says that the main issue is what happens to indigenous communities as development moves forward. Communities that have secured title to their lands can gain access to health and education services through development. As oil and gas production expands, however, it will encroach on the uncontacted tribes, which have no formal rights or title to the land they live on. “It’s an issue of human rights,” says Gamboa.

What is clear is that the operations in southeast Peru are just the beginning. Pluspetrol has its eyes on indigenous territories to the east of Camisea, and other companies are moving into the region. The Spanish company Repsol is exploring to the north, and Brazil’s government-controlled energy giant Petrobras is working to the west and south. The Peruvian government has already granted exploration rights for roughly half of its Amazon territory, and Babbitt says that it is just a matter of time before oil and gas companies seek access to the crown jewel of the Peruvian Amazon, Manu National Park, the country’s first major conservation initiative and an area of extreme biodiversity.

In addition to funding Peruvian non-governmental organizations to conduct research and draft legislative proposals, Babbitt and Ortiz are busy spreading the word within industry and government. Babbitt spoke about the offshore–inland idea during a conference this autumn on Latin American energy investments in New York, and he

is now taking the message to investor groups that promote corporate responsibility. He and Ortiz are also expanding their audience within South America by initiating talks with the Brazilian Development Bank and the Brazilian industrial giant Odebrecht Group, based in Salvador, which is working on a big pipeline into the Camisea region with Petrobras, based in Rio de Janeiro.

Babbitt acknowledges that it is a tough sell, even among the companies that have developed Camisea and proudly tout its environmentally friendly attributes. Two years ago, the consortium running the pipeline sought to build a second pipeline, which was to be routed through the lower Urubamba River valley, and through an important environmental sanctuary.

That plan raised alarm because to construct a pipeline, a temporary road would need to be built to move equipment around. Settlers often follow roads, so temporary roads become permanent ones, leading to deforestation and expanded development. That was not a risk with the first pipeline because the tortuous path it follows up and down hills is not attractive to loggers, miners and others seeking new territory.

When Babbitt and Ortiz learned about the plans for a new pipeline, they immediately moved to block it. Ortiz mobilized — and in some cases funded — environmental groups in Peru while Babbitt lobbied old contacts on Capitol Hill and at the IDB. They also made their case to Thomas Lovejoy, an Amazon ecologist at George Mason University in Fairfax, Virginia, who is on the advisory board of Hunt Oil in Dallas, Texas — one of Pluspetrol’s partners in the Camisea project and in the pipeline consortium. Hunt Oil officials were sympathetic to Babbitt and Ortiz’s case and pushed for a new pipeline route. In the face of growing opposition from environmental and indigenous groups, the consortium this year redirected the pipeline away from the sanctuary and up into the hills.

Despite Amazon Watch’s reservations about Camisea, its executive director, Atossa Soltani, gives Babbitt and Ortiz credit for helping to

“THIS IS FAR FROM AN IDYLIC FOREST, BUT WHAT I SEE IS SOMETHING THAT IS CONTAINED.”

persuade the companies to rethink the pipeline. “They took a tactical approach,” she says, “and they were really effective.”

Officials at Pluspetrol, however, do not hold up Camisea as a model. Nelson Soto Fuentes, the firm’s director of environmental issues and community relations, warns that roadless development is both expensive and difficult, and that some communities might actually welcome roads.

Others are more open to the idea. Carlos del Solar, a consultant for Hunt Oil, says that companies are increasingly looking to this kind of development as a way to avoid the social conflicts that afflict many major infrastructure projects in the region. “It’s expensive, but in the long run it pays off,” del Solar says, although he is not sure that government regulation is necessary.

As he prepares to leave Malvinas, Babbitt puts Camisea in a historical context, comparing the Amazon of today to America’s lawless Wild West. Although the struggle between conservation and development continues across the West, including in Babbitt’s Arizona, the US government has imposed the rule of law and set up a process to manage public lands. Babbitt hopes that the same will happen in the Amazon.

His goal is to slow down the most destructive road building and development until that time comes. The offshore–inland model is not a complete answer, he says, “but it starts the process of thinking about development in the right way.” ■

Jeff Tollefson covers energy and environment for *Nature* in New York.

1. Perz, S. *et al.* *Phil. Trans. R. Soc. B* **363**, 1889–1895 (2008).
2. Finer, M., Jenkins, C. N., Pimm, S. L., Keane, B. & Ross, C. *PLoS One* **3**, e2932 (2008).
3. Soares-Filho, B. S. *et al.* *Nature* **440**, 520–523 (2006).



CELL ANATOMY

► BY ROBERTA KWOK

A menagerie of intriguing cell structures, some long-neglected and others newly discovered, is keeping biologists glued to their microscopes.

In 2008, Chalongrat Noree faced an unenviable task: manually surveying hundreds of yeast strains under a microscope. Each strain had a different protein tagged with a fluorescent label, and Noree, a graduate student at the University of California, San Diego, was looking for interesting structures in the cells.

But it wasn't long until Noree's labour yielded results: within a month, he began finding a wide variety of proteins assembling into clusters or long strands. "Imagine every week you found a new intracellular structure," says Jim Wilhelm, a cell biologist and Noree's adviser. "If it were a slot machine, it would be paying off every other time you pulled the handle."

These days, textbook diagrams of cell structures such as the nucleus, mitochondrion, ribosome and Golgi apparatus are beginning to seem out of date. New imaging techniques, genome data, interest from disciplines outside cell biology and a bit of serendipity are drawing attention to an intricate landscape of tubes, sacs, clumps, strands and capsules that may be involved in everything from intercellular communication to metabolic efficiency. Some could even be harnessed for use in drug delivery or in synthesis of industrial products, such as biofuels.

Some of these structures have been known for decades, whereas others have only recently come to light. Wilhelm's team, for instance, has found six kinds of filament that either had never been described, or had been largely passed over. "You figure, how many structures could have been missed in the cell?" says Wilhelm. "Apparently, a lot more than you would imagine."

► LINES OF COMMUNICATION

One structure that is receiving fresh scrutiny is the membrane nanotube: a thin thread of membrane suspended between cells. In 2000, Amin Rustom, then a graduate student at Heidelberg University in Germany, was using a newly acquired dye to look at rat tumour cells under a fluorescence microscope. But he decided to skip some washing steps in the protocol. "He said, 'I saw something — I don't know what it is, but it looks interesting,'" recalls his former adviser, Hans-Hermann Gerdes, a cell biologist now at the University of Bergen in Norway. The tubes that Rustom had noticed were so straight that Gerdes initially wondered if they were scratches on the dish.

The team concluded in a 2004 study¹ that the structures, which could span the distance of several cells, were channels that could transport small cellular organelles. That same year, Daniel Davis, a molecular immunologist at Imperial College London, and his colleagues proposed that immune cells might send signals to each other along such tubes². At the time, Davis recalls, "There would always be people in the audience who would say, 'I saw those strands in the late 1970s or 80s.'" But earlier observers paid little heed to the tubes.

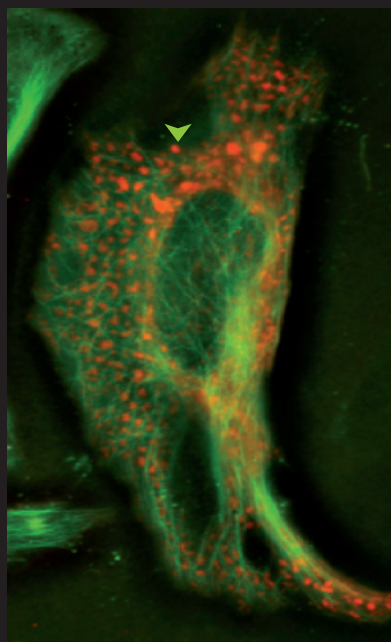
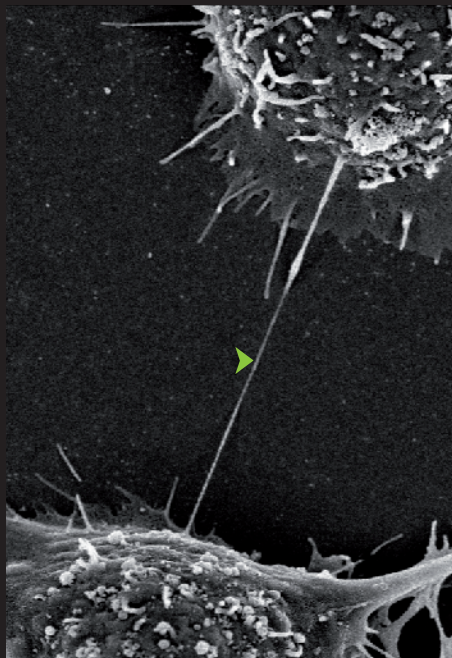
The 2004 reports prompted more studies, which have found nanotubes in many types of mammalian cell. Davis's team found that nanotubes could help certain white blood cells to kill cancer cells, either by acting as a tether that draws the cancer cell close or by providing a conduit for delivering lethal signals³. Nanotubes can also conduct electrical signals,

► NATURE.COM

For a podcast and
slideshow on this
story see:
go.nature.com/6elecp

► LINES OF COMMUNICATION

Nanotubes are thin membranous threads that run between cells and may serve as signalling conduits.

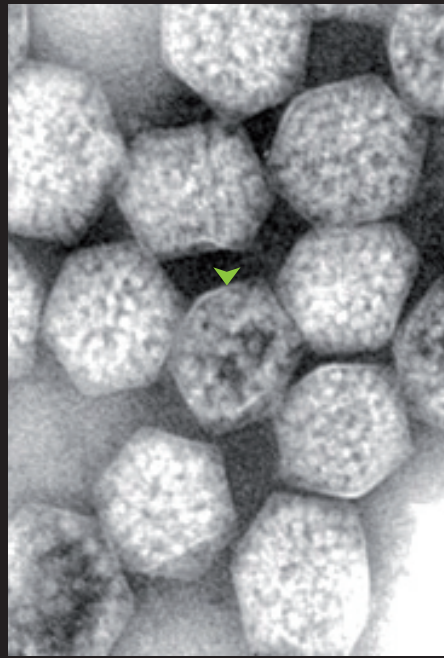


► PRODUCTIVITY HOTSPOTS

Some metabolic enzymes clump together under certain conditions, possibly to improve their efficiency.

► TINY FACTORIES

Orderly microcompartments seem to cordon off certain chemical reactions from the rest of the cell.



which might enable cells to coordinate during migration or wound healing, according to a 2010 study by Gerdes and his colleagues⁴. HIV and prions — infectious, misfolded proteins — may even travel along the tubes^{5,6}.

Some researchers are sceptical that nanotubes can form open channels. “It’s not clear that there’s a real continuous tunnel,” says Jennifer Lippincott-Schwartz, a cell biologist at the US National Institutes of Health in Bethesda, Maryland. And so far, nanotubes have been studied mainly in cell culture. Blocking nanotube formation in living organisms might give clues to their importance, says Davis. But such manipulations often disturb other crucial processes.

► PRODUCTIVITY HOTSPOTS

Researchers have long puzzled over how some metabolic processes work so efficiently. If the proteins involved are not close together, intermediate molecules could get lost in the “bewildering mass of enzymes in the cell”, says Stephen Benkovic, a chemical biologist at Pennsylvania State University in University Park. Proteins often assemble to carry out a particular task — a large complex is required to copy DNA, for example — but Benkovic and others have wondered whether metabolic enzymes might cluster together in a multistep assembly line, passing sometimes-unstable molecules from one ‘worker’ to the next.

Benkovic’s group found evidence that this clustering does occur in enzymes that produce a precursor of purine nucleotides, which are components of DNA and RNA. The team tagged each enzyme with a fluorescent label and observed them in living cells under the microscope. When a cell was deprived of purines, the enzymes grouped together in a cluster, which the team called the ‘purinosome’⁷. Last year, the team reported that purinosomes are nestled in a mesh of protein fibres called microtubules, like berries in a bramble bush⁸. The molecules produced by purinosomes can be converted to the cellular fuel adenosine triphosphate, so Benkovic speculates that purinosomes may help power the transport of organelles and materials around the cell on microtubule tracks.

Edward Marcotte, a systems biologist at the University of Texas at Austin, advises caution in interpreting these results, however. He and his colleagues have seen enzyme clusters as well: in 2009, they reported

that they had found 180 types of protein forming clumps in starved yeast cells⁹. But it is not clear whether the clumps serve a useful purpose — such as improving metabolic efficiency or acting as storage depots — or are a result of cellular failures brought on by starvation, says Marcotte.

► TINY FACTORIES

Some researchers are taking a closer look at elegant bacterial protein containers called microcompartments. First seen about 50 years ago, these polyhedron-shaped protein capsules resemble the outer shell of a virus¹⁰. But unlike viruses, which package genetic material, microcompartments contain enzymes that carry out important reactions, such as converting carbon dioxide into a form of carbon that is usable by the cell. Scientists suspect that the shells make reactions more efficient, keep toxic intermediate products away from the rest of the cell and protect enzymes from molecules that could hinder their performance.

In 2005, protein crystallographers helped to reveal the capsules’ finer details. Microcompartments “simply hadn’t attracted the attention yet of structural biologists”, says Todd Yeates, a structural biologist himself at the University of California, Los Angeles. He and his colleagues found that some shell proteins assemble into six-sided tiles that come together to form the sides of a microcompartment¹¹. Each tile has a hole in the centre that could allow molecules to pass through.

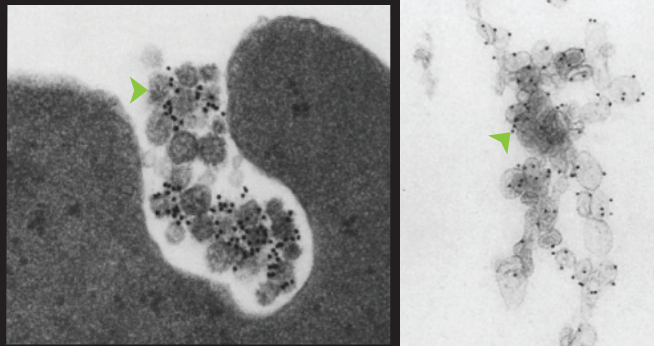
In addition to having an orderly structure, microcompartments can also line up in neat rows. Pamela Silver, a synthetic biologist at Harvard Medical School in Boston, Massachusetts, and her colleagues reported¹² last year that in cyanobacteria, certain microcompartments called carboxysomes “more or less stayed in a line down the centre of the cell”, says Silver. This tidy arrangement allows cells to allot carboxysomes evenly to daughter cells when dividing.

Biologists are now eager to exploit these capsules for industrial uses by loading them with different enzymes. For instance, Yeates and his team are planning to try engineering microcompartments to produce biofuel. Some researchers have managed to package fluorescent proteins or enzymes from other species into the shells, suggesting that it is possible to modify the capsules’ contents.

Microcompartments still offer plenty of unexplored territory. Scientists aren’t sure, for instance, exactly how enzymes are organized

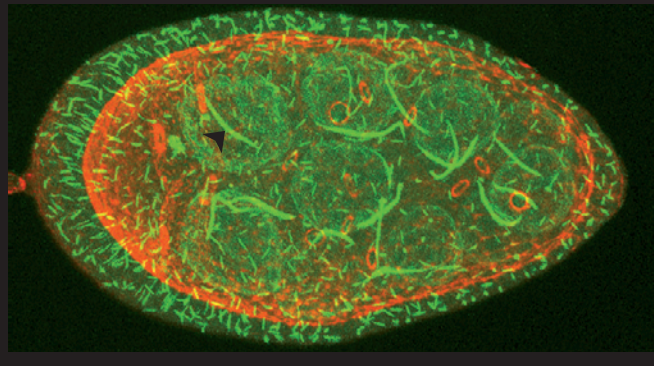
► CARGO CONTAINERS

Exosomes (seen here in two different environments) can parcel up and distribute molecules to distant cells.



► CELL SERPENTS

Some enzymes form filaments that serve as structural scaffolding and may control reactions en masse.



inside the capsules, says Cheryl Kerfeld, a structural biologist at Lawrence Berkeley National Laboratory in Berkeley, California. “We don’t really know what it looks like in there.”

► CARGO CONTAINERS

Other subcellular packages drawing attention are exosomes — tiny membrane-enclosed sacs that form inside the cell and are later spat out. These nanoscale vessels were discovered in the 1980s and then ignored for about a decade — considered a way of bagging up cellular rubbish. “People thought they were junk, basically,” says Jan Lötval, a clinical allergist at the University of Gothenburg in Sweden.

Interest in exosomes picked up in 1996, when Graça Raposo, a cell biologist now at the Curie Institute and the National Centre for Scientific Research in Paris, and her colleagues scrutinized exosomes spat out by B cells, a type of white blood cell. Although the technology to examine them — electron microscopy — wasn’t new, it wasn’t very popular at the time because “it was just old-fashioned”, says Raposo. Using it and other techniques, the team reported that the humble vessels might do something useful: display scraps of pathogen protein on their surfaces, spurring immune cells to mount defences against an infection¹³. Scientists became even more intrigued when Lötval’s team reported in 2007 that exosomes could carry messenger RNA¹⁴, some of which could be picked up and translated in a recipient cell. This suggested that the shipments might allow cells to affect protein production in their neighbours. The study “really showed that exosomes were a vehicle of communicating important information between cells”, says Clotilde Théry, a cell biologist who is also at the Curie Institute.

Researchers are now trying to use exosomes to deliver drugs to specific parts of the body — with the hope that, because exosomes are ‘natural’, they might be less likely to be toxic or provoke an immune response

than other vessels, such as artificial lipid sacs or protein shells. This year, Matthew Wood, a neuroscientist at the University of Oxford, UK, and his colleagues reported¹⁵ an attempt in mice: the team loaded exosomes with artificial RNA intended to hinder production of a protein involved in Alzheimer’s disease and tagged them with a molecule directing them to neurons and the blood–brain barrier. The exosomes successfully delivered their cargo and reduced production of the protein with no obvious ill effects, the team found. Other scientists are trying to fish exosomes out of body fluids and analyse their contents to diagnose cancer or deploy exosomes to provoke immune responses against tumours.

► CELL SERPENTS

Finally, Wilhelm’s group and others have found filaments that string together enzymes by the hundreds or thousands — enough, in some cases, to span nearly the entire cell. One of the filament-forming enzymes Wilhelm’s team found was CTP synthase, which makes a building block for DNA and RNA¹⁶. Two other teams discovered the same filaments in fruitflies and bacteria at around the same time^{17,18}. One researcher, Ji-Long Liu, a cell biologist at the Medical Research Council Functional Genomics Unit at the University of Oxford, named them cytoophidia (or ‘cell serpents’) because of their snake-like shapes in fly cells. Wilhelm suspects that researchers found the same filaments in the 1980s but never identified the protein.

These structures could allow the cell to turn enzymes on and off en masse, suggests Wilhelm. For instance, if the enzymes in a filament are inactive, the cell could activate all of them by dissolving the strand.

In some bacteria, enzyme filaments also seem to serve a structural purpose, somewhat like the actin filaments that are part of the cytoskeleton in more complex cells. When Zemer Gitai, a cell biologist at Princeton University in New Jersey, and his colleagues studied the structures in a comma-shaped bacterium called *Caulobacter crescentus*, they found that CTP-synthase filaments kept the cells’ curvature in check. If there was too little of the enzyme, the cells curled up tightly; if there was too much, they straightened out¹⁸.

It is not clear why curvature is important for the bacterium, says Gitai, but the findings suggest that the cells may have co-opted enzyme filaments to preserve cell shape. Researchers already suspect that actin is related to the enzyme hexokinase. It is possible that the cytoskeleton arose from filaments that originally formed to regulate the cell’s metabolism, Gitai says.

Although the purpose and importance of some of these emerging structures is not yet clear, the research illustrates that the act of simply observing cells and their contents is alive and well. “A key aspect of doing great science is exploration,” says Davis. “I think that there’s a tremendous amount that we learn just by watching.”■

Roberta Kwok is a freelance writer in Burlingame, California.

1. Rustom, A., Saffrich, R., Markovic, I., Walther, P. & Gerdes, H.-H. *Science* **303**, 1007–1010 (2004).
2. Önfelt, B., Nedvetzki, S., Yanagi, K. & Davis, D. M. *J. Immunol.* **173**, 1511–1513 (2004).
3. Chauveau, A., Aucher, A., Eissmann, P., Vivier, E. & Davis, D. M. *Proc. Natl Acad. Sci. USA* **107**, 5545–5550 (2010).
4. Wang, X., Veruki, M. L., Bukoreshtliev, N. V., Hartveit, E. & Gerdes, H.-H. *Proc. Natl Acad. Sci. USA* **107**, 17194–17199 (2010).
5. Sowinski, S. et al. *Nature Cell Biol.* **10**, 211–219 (2008).
6. Gousset, K. et al. *Nature Cell Biol.* **11**, 328–336 (2009).
7. An, S., Kumar, R., Sheets, E. D. & Benkovic, S. J. *Science* **320**, 103–106 (2008).
8. An, S. et al. *Proc. Natl Acad. Sci. USA* **107**, 12872–12876 (2010).
9. Narayanaswamy, R. et al. *Proc. Natl Acad. Sci. USA* **106**, 10147–10152 (2009).
10. Drews, G. & Niklowitz, W. *Arch. Mikrobiol.* **24**, 147–162 (1956).
11. Kerfeld, C. A. et al. *Science* **309**, 936–938 (2005).
12. Savage, D. F., Afonso, B., Chen, A. H. & Silver, P. A. *Science* **327**, 1258–1261 (2010).
13. Raposo, G. et al. *J. Exp. Med.* **183**, 1161–1172 (1996).
14. Valadi, H. et al. *Nature Cell Biol.* **9**, 654–659 (2007).
15. Alvarez-Erviti, L. et al. *Nature Biotechnol.* **29**, 341–345 (2011).
16. Noree, C., Sato, B. K., Broyer, R. M. & Wilhelm, J. E. *J. Cell Biol.* **190**, 541–551 (2010).
17. Liu, J.-L. *J. Genet. Genomics* **37**, 281–296 (2010).
18. Ingerson-Mahar, M., Briegel, A., Werner, J., Jensen, G. & Gitai, Z. *Nature Cell Biol.* **12**, 739–746 (2010).

COMMENT

CLIMATE CHANGE Experts fear permafrost thaw will release more carbon than thought **p.32**

CELL BIOLOGY Time to ditch HeLa lines and move to stem cells **p.34**

NUCLEAR SCIENCE Cold-war defection of an Italian physicist **p.35**



OBITUARY John McCarthy, the father of artificial intelligence **p.40**

H. G. PONTING/POPPERFOTO/GETTY



Terra Nova: Robert Scott's ill-fated 1911 expedition opened up the Antarctic to scientific exploration.

Turning the world upside down

Research, not pole-bagging, was the lasting achievement of Antarctic exploration 100 years ago, says **Edward J. Larson**.

This month, and again in January, hundreds of scientists across Antarctica will set down their tools to mark the 100th anniversaries of the first explorers reaching the South Pole. Most will see the centenaries as simply marking the end of a much-romanticized race between Roald Amundsen's dog-sledding Norwegians and Robert Scott's man-hauling Brits. Yet they should also celebrate those journeys, and the competition involved, for their role in launching Antarctic science.

Competition drove early Antarctic research much as it still drives modern science, but the contest did not begin with Amundsen and Scott's 1911 race to the pole. It had started a decade earlier in the rivalry between Britain and Germany to discover a continent for science. With Britain's Royal Society and Royal Geographic Society and German geophysicist Georg von Neumayer leading the way, a generation of scientists in both countries had lobbied their governments for big-science projects

in the Antarctic. In 1900, scientists in each country used the threat of the other nation gaining the advantage in polar discovery to prod their own to fund what were planned as the first expeditions to winter in the Antarctic. Others rushed in with less ambitious ventures, but the expeditions aboard purpose-built research ships — Britain's *Discovery* and Germany's *Gauss* — launched the heroic age of Antarctic exploration.

What was the science being sought? In 1900, no one knew if Antarctica was ►



Robert Scott, Ernest Shackleton and Edward Wilson (left to right) setting out on the first attempt on the South Pole in 1902.

▶ a continent or simply a polar ice cap anchored on an archipelago of islands. Yet studies of deep-sea currents and high-altitude weather patterns suggested that the Antarctic governed ocean circulation and influenced the global climate. On the basis of seabed deposits and the fossil record further north, some scientists argued that it must contain a continental landmass, which formerly had been warm enough to support higher plants and had been linked to the other southern continents, serving as a land bridge for species movement. Charles Darwin had proposed the latter hypothesis to account for the abrupt appearance of imprints of *Glossopteris* plants in the fossil record of Africa, South America and Australia. Geophysicists sought reliable data on terrestrial magnetism for deep-southern latitudes and wanted to determine the precise location of the South Magnetic Pole, to which the compass points — a slowly migrating region that, at the time, was located nearly 2,000 kilometres north of the South Geographic Pole.

FIRST STEPS

Commanded by geophysicist Erich von Drygalski, the *Gauss* left Germany in the summer of 1901. It became trapped in pack ice off the little-known Antarctic coast south of the Indian Ocean and its scientists made only limited excursions onshore. They conducted oceanographic, magnetic, meteorological and wildlife research from the vessel for nearly a year before getting out of the ice in February 1903 and heading home prematurely.

Under Scott's command, the *Discovery* expedition also sailed south in 1901. The British team established its winter quarters

on Ross Island in a region already known to contain a coastal mountain range, an active volcano, rich marine life and an ice shelf that some thought might extend to the South Pole. While researchers collected marine and geological specimens, took continual seismic, magnetic and meteorological readings, searched for fossils, discovered the first-known emperor penguin rookery, measured the movement of glaciers and mapped the terrain and coastline, Scott set off for the pole with ship's officer Ernest Shackleton and surgeon Edward Wilson. Although they fell far short of their goal because of poor planning and execution, they marched farther south than any previous party and transformed reaching the pole into a British obsession. Over the next decade, Shackleton and Scott returned to the region with two new expeditions — Shackleton's *Nimrod* expedition (1907–09) and Scott's ill-fated *Terra Nova* expedition (1910–13) — that featured both a polar dash and scientific research.

At a time when the British Empire ruled one-quarter of the world's land area and sought to extend its reach still farther, these British expeditions captured international attention. Newspapers, magazines and book publishers vied for first-hand accounts by Scott and Shackleton. Manufacturers of products ranging from oatmeal and cigarettes to scientific equipment and motor cars supplied their goods free to the explorers in exchange for the chance to tout their use in Antarctica.

Scientists lined up to serve, with Shackleton snagging the distinguished geologist

Edgeworth David as scientific director for his *Nimrod* expedition. Accompanied by his former student Douglas Mawson and others, David climbed the southernmost active volcano and man-hauled heavy sledges more than 1,000 kilometres across sea ice, up a glacier and over the Antarctic Plateau to reach the South Magnetic Pole in January 1909. Such extreme research efforts were typical of the British expeditions.

Stung by Scott's criticism of his physical shortcomings — supposedly a lack of endurance — on *Discovery*'s polar trek, Shackleton had returned south in 1907 with the clear goal of reaching the geographic pole, even as David and the other scientists on his expedition pursued a rich research agenda. Besting the researchers on the *Discovery* expedition on virtually every score, they determined that Antarctica is a continental landmass, plotted the atmospheric patterns radiating from the South Polar Region, found widespread evidence of retreating glaciers, discovered freshwater algae and microorganisms in frozen lakes, and obtained fossil evidence of higher plant life in a land now devoid of it. Antarctica had experienced a temperate climate in earlier epochs, they concluded, and was now warming again. Meanwhile, displaying the determination and survival skills that would make him a legend, Shackleton pioneered a route south across the ice shelf, up the Beardmore Glacier and over the Antarctic Plateau to within 200 kilometres of the South Geographic Pole. It earned him a knighthood on his return home.

SCOTT RETURNS

With the pole still unclaimed, and intent on restoring his pre-eminence in Antarctic exploration, Scott headed south in 1910 aboard the ship *Terra Nova* with more scientists than on any previous polar expedition. "Doubtless there are those who will criticize this provision [for scientific research] in view of its published objectives — that of reaching the South Pole," Scott wrote of his final expedition. "But I believe that the more intelligent section of the community will heartily approve of the endeavour to achieve the greatest possible scientific harvest which the circumstances permit." Scott received more than 8,000 applications for 32 places on the shore party.

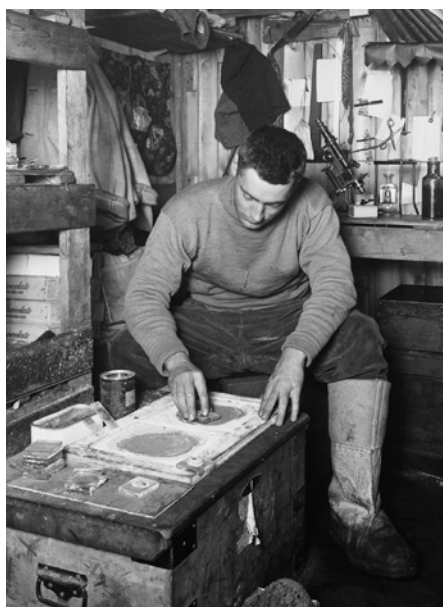
Not knowing of Amundsen's intentions before setting sail, Scott planned a deliberate assault on the pole. Like Amundsen, Scott wintered on the edge of the Ross Sea before embarking in the spring on the arduous trek to the pole. But unlike Amundsen, who travelled light and only with those who would go the entire 1,200-kilometre distance, Scott set off with teams of men, ponies, tractors and dogs that would leave depots of supplies and fall back in stages until only one group remained to man-haul a single sledge over

➔ **NATURE.COM**
For a podcast
on this story,
go to:
go.nature.com/ytfpwk

H. G. PONTING/POPPERFOTO/GETTY



H. G. PONTING/SPRUI, UNIV. CAMBRIDGE (LEFT); H. G. PONTING/POPPERFOTO/GETTY



Terra Nova science (clockwise from top): photographer H. G. Ponting in the Adélie penguin rookery at Cape Royds in 1911; inflating a weather balloon; geologist Frank Debenham grinds rock samples.

the Antarctic Plateau. Scott designed the mission for safety, not speed, and was unable or unwilling to adjust it after learning that Amundsen was also heading south.

Scott's expedition was further burdened with a weighty research agenda. During the first winter, while the Norwegians focused on preparing for their polar dash, the British engaged in a variety of activities. Six men, including geologist Raymond Priestly, a veteran of the Nimrod expedition with proven survival skills, had been dispatched to survey the Ross Sea's northwestern coast, where they spent the winter. The following summer, rather than returning this party to the main base, Scott transferred it to a geologically interesting site farther south on the coast for

a planned month of research, which extended over a second winter when ice blocked the *Terra Nova* ship from extracting the group. Although it made notable findings, the party contributed nothing to the polar dash.

Edward Wilson, the *Terra Nova*'s scientific director, caused another costly diversion by taking a 225-kilometre winter journey to an Emperor penguin rookery. Having visited the site during the Discovery expedition, he wanted to study the birds during the midwinter nesting period. No one had ever tried to travel overland in the dark Antarctic winter, much less haul sledges at temperatures down to -60°C . It nearly killed the three men involved, but two of them, Wilson and H. R. Bowers,

subsequently accompanied Scott to the pole.

A full regime of research continued through the Antarctic winter and into the spring. A month before setting off on the polar journey, Scott himself led a 240-kilometre trek into the Trans-Antarctic Mountains to measure the movement of the Ferrar Glacier. Critics have chided Scott for such actions, which inevitably handicapped him in the race with Amundsen. Two days before departing for Ferrar, however, after reviewing the expedition's scientific work and weighing it against the likelihood of Amundsen beating him to the pole, Scott wrote in his diary: "It is really a satisfactory state of affairs all around. If the [polar] journey comes off, nothing, not even lost priority at the Pole, can prevent the Expedition ranking as one of the most important that ever entered the Polar regions." Science would make it so, and so science must be served.

Of course, the polar journey did not succeed. The Norwegians beat the British to the South Pole and Scott and his four companions died on the way back. The weather was unseasonably cold and the men steadily weakened. Nevertheless, near the end, they stopped to collect fossils from beside the Beardmore Glacier. Wilson had spotted *Glossopteris* impressions, which geologists on all the British expeditions had sought in hopes of proving that the southern continents were once linked to Antarctica. Scott's party hauled these specimens to the final camp, where they were later found with the dead men, their diaries and their field notes.

By the time the findings from the *Terra Nova* expedition had been fully analysed back in Britain, scientists had gained a new appreciation of the Antarctic's global significance. Previously, European and US Earth scientists had given priority to the Northern Hemisphere. With the voyage of the British research ship *Challenger* in the 1870s, they began to realize that undercurrents from the Southern Ocean regulate the marine life and the temperature of the other great oceans, and that air currents from the far south influence the climate farther north. Data collected by Antarctic explorers during the early twentieth century, in large part by scientists travelling with Shackleton and Scott, confirmed that Antarctica's vast glaciers, ice sheets and ice shelves — much more than the Arctic's thin ice cap — influence the global environment. In doing so, they turned the world upside down and put Antarctica on top. ■

Edward J. Larson is university professor of history and law at Pepperdine University, Malibu, California 90263, USA. His most recent book is *An Empire of Ice: Scott, Shackleton, and the Heroic Age of Antarctic Science* (Yale Univ. Press, 2011). e-mail: ed.larson@pepperdine.edu



Abrupt thaw, as seen here in Alaska's Noatak National Preserve, causes the land to collapse, accelerating permafrost degradation and carbon release.

High risk of permafrost thaw

Northern soils will release huge amounts of carbon in a warmer world, say
Edward A. G. Schuur, Benjamin Abbott and the Permafrost Carbon Network.

Arctic temperatures are rising fast, and permafrost is thawing. Carbon released into the atmosphere from permafrost soils will accelerate climate change, but the magnitude of this effect remains highly uncertain. Our collective estimate is that carbon will be released more quickly than models suggest, and at levels that are cause for serious concern.

We calculate that permafrost thaw will release the same order of magnitude of carbon as deforestation if current rates of deforestation continue. But because these emissions include significant quantities of methane, the overall effect on climate could be 2.5 times larger.

Recent years have brought reports from the far north of tundra fires¹, the release of ancient carbon², CH₄ bubbling out of lakes³ and gigantic stores of frozen soil carbon⁴. The latest estimate is that some 18.8 million square kilometres of northern soils hold about 1,700 billion tonnes of organic carbon⁴ — the remains of plants and animals that have been accumulating in the soil over thousands of years. That is about four times more than all the carbon emitted by human activity in modern times and twice as much as is present in the atmosphere now.

This soil carbon amount is more than three times higher than previous estimates,

largely because of the realization that organic carbon is stored much deeper in frozen soils than was thought. Inventories typically measure carbon in the top metre of soil. But the physical mixing during freeze–thaw cycles, in combination with sediment deposition over hundreds and thousands of years, has buried permafrost carbon many metres deep.

The answers to three key questions will determine the extent to which the emission of this carbon will affect climate change: How much is vulnerable to release into the atmosphere? In what form it will be released? And how fast will it be released? These questions are easily framed, but challenging to answer.

KNOWN UNKNOWNs

As soils defrost, microbes decompose the ancient carbon and release CH₄ and carbon dioxide. Not all carbon is equally vulnerable to release: some soil carbon is easily metabolized and transformed to gas, but more complex molecules are harder to break down. The bulk of permafrost carbon will be released slowly over decades after thaw, but a smaller fraction could remain within the soil for centuries or longer. The type of gas released also affects the heat-trapping potential of the emissions. Waterlogged, low-oxygen environments are likely to contain microbes that produce CH₄ — a potent

greenhouse gas with about 25 times more warming potential than CO₂ over a 100-year period. However, waterlogged environments also tend to retain more carbon within the soil. It is not yet understood how these factors will act together to affect future climate.

The ability to project how much carbon will be released is hampered both by the fact that models do not account for some important processes, and by a lack of data to inform the models. For example, most large-scale models project that permafrost warming depends on how much the air is warming above them. This warming then boosts microbial activity and carbon release. But this is a simplification. Abrupt thaw processes can cause ice wedges to melt and the ground surface to collapse, accelerating the thaw of frozen ground⁵. Evidence of rapid thaw is widespread: you can see it in the 'drunken' trees that tip dangerously as a result of ground subsidence, and in collapsed hill slopes marked by scars from landslides. These are just some of the complex processes that models don't include.

At the same time, few data are available to support these models because of the difficulties of gathering data in extreme environments. Only a handful of remote field stations around the world are collecting data to support this research, even though the permafrost zone covers about almost one-quarter

of the Northern Hemisphere's land area. The field studies that do exist confirm that permafrost thaw is tightly linked to ground subsidence and soil moisture as well as temperature. So modelling carbon emissions from permafrost thaw is much more complex than a simple response to temperature alone.

Models have flaws, but experts intimately familiar with these landscapes and processes have accumulated knowledge about what they expect to happen, based on quantitative data and qualitative understanding of these systems. We have attempted to quantify this expertise through a survey developed over several years.

SURVEY SAYS

Our survey asks what percentage of the surface permafrost is likely to thaw, how much carbon will be released, and how much of that carbon will be CH_4 , for three time periods and under four warming scenarios that will be part of the Intergovernmental Panel on Climate Change Fifth Assessment Report. The lowest warming scenario projects 1.5 °C Arctic warming over the 1985–2004 average by the year 2040, ramping up to 2 °C by 2100; the highest warming scenario considers 2.5 °C by 2040, and 7.5 °C by 2100. In all cases, we posited that the temperature would remain steady from 2100 to 2300 so that we could assess opinions about the time lag in the response of permafrost carbon to temperature change.

The survey was filled out this year by 41 international scientists, listed as authors here, who publish on various aspects of permafrost. The results are striking. Collectively, we hypothesize that the high warming scenario will degrade 9–15% of the top 3 metres of permafrost by 2040, increasing to 47–61% by 2100 and 67–79% by 2300 (these ranges are the 95% confidence intervals around the group's mean estimate). The estimated carbon release from this degradation is 30 billion to 63 billion tonnes of carbon by 2040, reaching 232 billion to 380 billion tonnes by 2100 and 549 billion to 865 billion tonnes by 2300. These values, expressed in CO_2 equivalents, combine the effect of carbon released as both CO_2 and as CH_4 .

Our estimate for the amount of carbon released by 2100 is 1.7–5.2 times larger than those reported in several recent modelling studies^{6–8}, all of which used a similar warming scenario. This reflects, in part, our perceived importance of the abrupt thaw processes, as well as our heightened awareness of deep carbon pools. Active research is aimed at incorporating these main issues, along with others, into models.

Are our projected rapid changes to the permafrost soil carbon pool plausible? The survey predicts a 7–11% drop in the size of the permafrost carbon pool by 2100 under the high-warming scenario. That scale of

carbon loss has happened before: a 7–14% decrease has been measured in soil carbon inventories across thousands of sites in the temperate-zone United Kingdom as a result of climate change⁹. Also, data scaled up from a single permafrost field site point to a potential 5% loss over a century as a result of widespread permafrost thaw². These field results generally agree with the collective carbon-loss projection made by this survey, so it should indeed be plausible.

Across all the warming scenarios, we project that most of the released carbon will be in the form of CO_2 , with only about 2.7% in the form of CH_4 . However, because CH_4 has a higher global-warming potential, almost half the effect of future permafrost-zone carbon emissions on climate forcing is likely to be from CH_4 . That is roughly consistent with the tens of billions of tonnes of CH_4 thought to have come from oxygen-limited environments in northern ecosystems after the end of the last glacial period¹⁰.

All this points towards significant carbon releases from permafrost-zone soils over policy-relevant timescales. It also highlights important lags whereby permafrost degradation and carbon emissions are expected to continue for decades or centuries after global temperatures stabilize at new, higher levels. Of course, temperatures might not reach such high levels. Our group's estimate for carbon release under the lowest warming scenario, although still quite sizeable, is about one-third of that predicted under the strongest warming scenario.

Knowing how much carbon will be released from the permafrost zone in this century and beyond is crucial for determining the appropriate response. But despite the massive amount of carbon in permafrost

soils, emissions from these soils are unlikely to overshadow those from the burning of fossil fuels, which will continue to be the main source of climate forcing. Permafrost carbon release will still be an important amplifier of climate change, however, and is in some ways more problematic: it occurs in remote places, far from human influence, and is dispersed across the landscape. Trapping carbon emissions at the source — as one might do at power plants — is not an option. And once the soils thaw, emissions are likely to continue for decades, or even centuries.

The scientific community needs to collect more data and develop more-sophisticated models to test the hypotheses presented by this survey. Fortunately, awareness of the problem is increasing and these are starting to happen. The US Department of Energy, for example, has initiated a project called Next-Generation Ecosystem Experiments — Arctic, which aims to improve the representation of these processes in large-scale models. NASA is pursuing an Arctic–Boreal Vulnerability Experiment, which aims to improve satellite observations of this region. The Vulnerability of Permafrost Carbon Research Coordination Network funded by the US National Science Foundation, of which we are part, is bringing together people and observations to synthesize results and validate models. These are just some of the many international initiatives aimed at filling these research gaps.

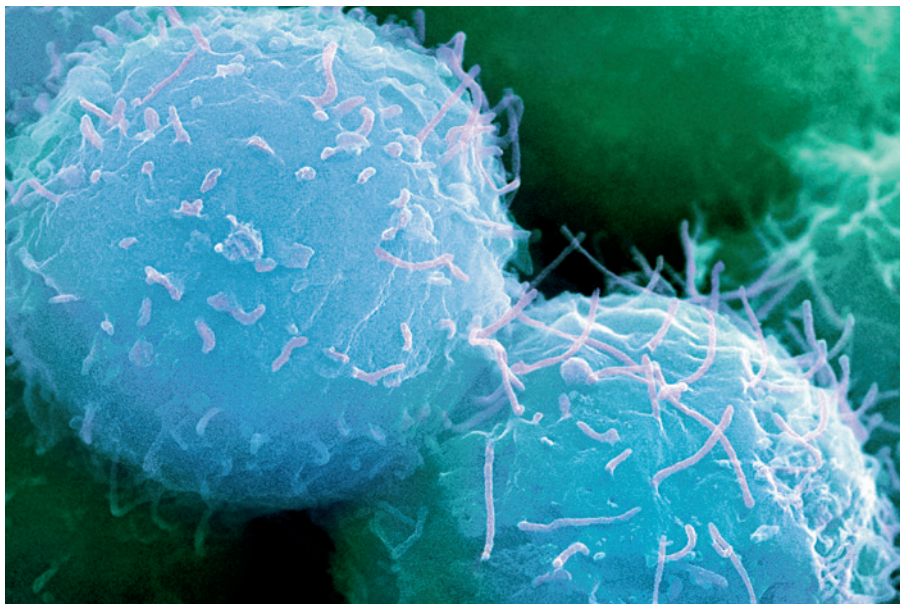
In the meantime, our survey outlines the additional risk to society caused by thawing of the frozen north, and underscores the urgent need to reduce atmospheric emissions from fossil-fuel use and deforestation. This will help to keep permafrost carbon frozen in the ground. ■

Edward Schuur is in the Department of Biology at the University of Florida, Gainesville, Florida 32611, USA. **Benjamin Abbott** is in the Institute of Arctic Biology at the University of Alaska, Fairbanks, Alaska 99775, USA. The other experts in the Permafrost Carbon Research Network are listed at go.nature.com/kkidom. e-mail: tschuur@ufl.edu



'Drunken' trees reveal areas of subsidence.

1. Mack, M. C. *et al.* *Nature* **475**, 489–492 (2011).
2. Schuur, E. A. G. *et al.* *Nature* **459**, 556–559 (2009).
3. Walter, K. M., Zimov, S. A., Chanton, J. P., Verbyla, D. & Chapin, F. S. III *Nature* **443**, 71–75 (2006).
4. Tarnocai, C. *et al.* *Global Biogeochem. Cycles* **23**, GB2023 (2009).
5. Jorgenson, M. T., Shur, Y. L. & Pullman, E. R. *Geophys. Res. Lett.* **33**, L02503 (2006).
6. Schaefer, K., Zhang, T., Bruhwiler, L. & Barrett, A. P. *Tellus B* **63**, 165–180 (2011).
7. Koven, C. D. *et al.* *Proc. Natl Acad. Sci. USA* **108**, 14769–14774 (2011).
8. Schneider von Deimling, T. *et al.* *Biogeosciences Discuss.* **8**, 4727–4761 (2011).
9. Bellamy, P. H., Loveland, P. J., Bradley, R. I., Lark, R. M. & Kirk, G. J. *Nature* **437**, 245–248 (2005).
10. Fischer, H. *et al.* *Nature* **452**, 864–867 (2008).



Stem cells offer biologists the chance to unpick the molecular differences that define cell types.

Beyond HeLa cells

To find out what distinguishes one cell type from another, cell biologists must renounce popular cell lines, argue **Anthony H. Hyman** and **Kai Simons**.

In 1896, E. B. Wilson defined the cell as the basis of the life of all organisms. Subsequent discoveries in enzymology and molecular biology demonstrated the unity of life: the basic machinery is similar in all cells, giving rise to Jacques Monod's famous dictum that what is true for *Escherichia coli* is also true for an elephant.

This understanding would have been impossible without the use of single-celled organisms and animal and human cell lines that will grow indefinitely in culture. The most famous of these is the continuously dividing HeLa line, derived from the cervical cancer cells of a woman named Henrietta Lacks. More than 65,000 scientific studies using HeLa cells have been published since the 1950s, and the cells have been used to study every conceivable aspect of cell physiology as well as the basic machinery common to all cells.

However, cancer cells, such as those from which the HeLa line was derived, have a physiology far removed from that of normal cells in tissue, where cells interact with and receive signals from other cells nearby. Furthermore, HeLa cells cannot tell researchers anything about what distinguishes one cell type from another — a liver cell from a pancreatic cell, for instance. So, although HeLa cells and other immortal cells derived from cancer patients are good

for investigating what cells have in common, they are completely inadequate for addressing the next big topic in cell biology: cellular diversity in normal tissue. In other words, how do cells' DNA, RNA and proteins act together to specify the properties of different cell types? To answer this question, we need to make the difficult transition to a new research tool in cell biology: stem cells.

Cytologists can look at a cell under the microscope and give it a label, but they do so on the grounds of its morphology — its external shape and the appearance of its internal structures, for instance. On this basis, we assign labels such as 'pyramidal cells', 'medium spiny neurons' or 'cuboidal epithelial cells'. But this is a crude approach, akin to how microbial species were defined before the invention of DNA sequencing. What really distinguishes one cell type from another is how the various molecular elements function and interact.

Understanding the molecular basis of cell identity has important medical implications. Pathologists generally define the type of cancer by cell morphology. But cancer genomics projects are showing that each type of cancer is very diverse: patterns of mutation in tumours of the same organ have little in common in different individuals. The literature is full of cases in which, for example, scientists believed they

were studying a breast-cancer cell line that turned out to be derived from melanoma cells instead. Only a molecular understanding of why cells look and behave differently will allow unambiguous definition of both normal and altered cell types.

One way to achieve this is to move away from immortalized tissue-culture cells; the days of HeLa cells are over. Instead, cell biologists should use embryonic stem cells derived from mice or other model organisms, or convert differentiated cells into precursors using a cocktail of transcription factors, and study in detail how these cells become various cell types.

At the moment, stem-cell research is heavily driven by the hope of new therapies that can replace malfunctioning cells in our body. To expand the field, cell and stem-cell biologists need to work together to derive, maintain and differentiate cell lines representing different cell types, and to use modern imaging and other methods in the cell biologist's tool-kit to study these cells. Just as developmental biology was transformed by insights and approaches from cell biology, it will be crucial to introduce those same approaches into stem-cell biology.

The movement of cell biologists into stem-cell biology will also aid attempts to harness the healing potential of stem cells in medicine. Currently, stem-cell research relies heavily on transcription factors to differentiate and to identify cells, because different cell types express different transcription factors. But this characterization is too narrow; cell identity encompasses more than a handful of transcriptional networks. If cell biologists could uncover the molecular machinery that distinguishes different cell types, this would make stem-cell research safer. To avoid the catastrophes of the early days of gene therapy, in which genetic modifications induced cancer, we must understand the cell biology of the cells that will be reintroduced. Do they really function as normal counterparts in our body? Or could they potentially turn into cancer cells in their new environment?

A switch from cell lines such as HeLa, which are simple to grow and maintain, to stem cells, which require more precise conditions, will be difficult. Consequently, funding organizations must develop standardized protocols and other tools that make it easier to produce the plethora of different cells that we need to study. In the 1960s, agencies dedicated funds to establish the first techniques that maintained cells in culture — it is time for a new technological breakthrough. ■

Anthony H. Hyman and **Kai Simons** are at the Max Planck Institute of Molecular Cell Biology and Genetics, 01307 Dresden, Germany.
e-mail: hyman@mpi-cbg.de



Italian nuclear physicist Bruno Pontecorvo fled to Moscow in 1950, but there is no evidence he was a spy.

PHYSICS

Atomic secrets

Mystery lingers round the sudden defection of cold-war physicist Bruno Pontecorvo, finds **Sharon Weinberger**.

Governments love secrets. Nuclear secrets are the most prized of all, and their potential revelation is greatly feared. That is why, more than 60 years after the event, the defection of Italian nuclear physicist Bruno Pontecorvo to the Soviet Union is still so unsettling.

Was Pontecorvo — the only Western scientist involved in wartime nuclear projects to have defected — a Soviet spy or an idealist seeking to escape anti-communist hysteria? In *The Pontecorvo Affair*, historian Simone Turchetti re-examines this intriguing incident.

Stymied by the secrecy still shrouding the case, Turchetti cannot tell us whether Pontecorvo was a Soviet agent or not. But he upends the notion that Pontecorvo's limited access to atomic 'secrets' made his defection a minor

footnote to cold-war history. He argues compellingly that it was Pontecorvo's expertise in key areas of nuclear physics, rather than access to secret work, that made his defection significant. "Scientific knowledge," Turchetti writes, "cannot be smuggled in plastic bags."

He also offers new insight into what impelled Pontecorvo to flee to the East in 1950, and notes that both the US and UK governments were eager to downplay Pontecorvo's nuclear expertise to minimize the public impact of his defection. Even if he was not an atomic spy, Pontecorvo probably made a meaningful contribution to the Soviet nuclear programme, particularly in the area of geophysical prospecting — something that has not been

NATURE.COM
For more on Bruno
Pontecorvo's work:
go.nature.com/oa6key



The Pontecorvo Affair: A Cold War Defection and Nuclear Physics
SIMONE TURCHETTI
University Of Chicago Press: 2012. 272 pp.
\$45, £29

appreciated before.

The story of Pontecorvo's career begins with the familiar narrative of a Second World War émigré. As a scientist from an Italian Jewish family, he fled the march of fascism by leaving for the United States in 1940. Although never directly involved in the Manhattan Project, he became one of several prominent Italian scientists to work on nuclear programmes, including Enrico Fermi and Emilio Segrè. His contributions to geophysical prospecting involved looking at the interaction of neutrons with rock formations, which help to reveal the presence of oil underground. He was also involved in the 'pile physics' of nuclear reactors.

His Italian nationality aroused suspicion among Western security services, however. His family was known for its communist leanings: his brother Gillo later directed the iconic anti-colonial 1966 film *The Battle of Algiers*.

For a while, the suspicions did not hamper Pontecorvo's work, even on sensitive nuclear projects. After a stint in Canada, Pontecorvo left for the United Kingdom, where he worked on the British nuclear-bomb project. Eventually, however, the security noose tightened and Pontecorvo was forced out of his job at the Atomic Energy Research Establishment in Harwell. In 1950, he accepted an academic position at Liverpool University. Then, the well-worn narrative veers off course. Midway through a seemingly bucolic summer holiday, Pontecorvo and his family began a sudden dash across Europe, eventually heading to Stockholm and on through Finland to the Soviet Union.

Little solid evidence has been offered for why Pontecorvo fled when he did. Some accounts allege that he was a Soviet agent, whisked away by handlers who feared his exposure — particularly after the conviction earlier that year of atomic spy Klaus Fuchs. Turchetti, however, argues that the trigger was the battle brewing over the slow-neutron patent filed by Pontecorvo and other Italian scientists. This gave them a financial stake in the technology the United States was using to produce plutonium. At the time, the US government was producing it in a reactor that used graphite to slow the neutrons.

In 1950, one of the patent holders, Gabriel Giannini, filed a suit against the US government, seeking compensation. The suit caught Pontecorvo off guard, and Turchetti argues that the prospect of having his name tied to a public legal battle against the

► US government, combined with the mounting security investigation into his background, was what prompted him to flee. The timing is persuasive: the day the news of the suit reached Europe, Pontecorvo set his plans for defection in motion.

Turchetti is critical of contemporaneous media accounts and later books that allege Pontecorvo was a spy; some are based on recollections of former agents of the Soviet security service. Decades on, there is no firm evidence to support the allegation.

What did Pontecorvo offer the Soviet atomic weapons programme? Here Turchetti speculates, based on what is known. After his defection, Pontecorvo worked at the Dubna Institute for Nuclear Research near Moscow (now the Joint Institute for Nuclear Research) — ostensibly only on civil nuclear science. Turchetti argues that, in reality, his expertise would have allowed him to make important contributions to the Soviet atomic bomb without necessarily working on secret projects. In particular, Pontecorvo's geophysical expertise may have helped the Soviet Union to gain access to uranium reserves, which were in short supply in the 1950s.

Today, neither the Russian archives nor those of the US Federal Bureau of Investigation on Pontecorvo are open. Such secrecy tells us much about those times and about the post-cold-war mindsets on both sides of the fence, which remain riddled with paranoia. The allegations of espionage levied in 1999 against Wen Ho Lee, a Taiwanese-US scientist working at Los Alamos National Laboratory in New Mexico, demonstrate how those fears continue to guide, and misguide, investigations of such claims.

Wen Ho Lee was exonerated of espionage. Pontecorvo — who died in Russia in 1993, deeply disillusioned with communism — remains in posthumous limbo.

One thing is clear. If Pontecorvo was not a spy, and his defection was based simply on fear of persecution and a preference for life under communism, he made a poor choice. In the West during the cold war, scientists lost their jobs amid witch-hunts, but in the Soviet Union, at least under Joseph Stalin, the outcomes were more dire. Nothing in the book illustrates this better than Stalin's comment on a proposed conference organized by nuclear physicists: "Leave them in peace, we can shoot them later." ■

Sharon Weinberger is a writer based in Washington DC and a Carnegie Fellow at the Medill School of Journalism, Northwestern University, Evanston, Illinois, USA.
e-mail: sharonweinberger@gmail.com



A dynamic eighteenth-century exhibit of horse muscle anatomy from Pavia's Museum of Natural History.

MUSEUMS

Stripped assets

Paolo Mazzarello argues that the disposal of collections requires clear consultation with the public.

Museums are facing a sort of Malthusian constraint — an explosive increase in the volume of their collections, coupled with a severe reduction in funding, fuelled partly by the current economic crisis. Collections cannot be increased indefinitely and sustained forever. So what happens when a saturation point is reached?

The Museum of Natural History at the University of Pavia in Italy is an example of how things can go wrong at such a tipping point. Founded by naturalist Lazzaro Spallanzani in the eighteenth century, the museum was dismantled in the 1930s in the face of financial and other pressures. The collection — including some of the first animal specimens preserved by taxidermy — was then dispersed across a number of sites in Pavia and the surrounding area.

This was an extreme solution, but museums must evolve. The goal of museums, to

safeguard the evidence of important changes in the history of the planet and humankind, is an endless task. In response, they must be dynamic places, where the acquisition of objects is balanced by the planned deaccession or disposal of others.

Such an activity is at odds with the fusty image of a museum as a place where items are preserved in display cases or kept in storage. However, the disposal of materials demands care: whether justifiable or not, it is often highly controversial and can devastate an institution's image. From small civic archives to the vast Prado in Madrid, museums are the repositories of our collective past and identity, and that makes any broad discussion of disposal problematic.

Careful accession and deaccession policies are becoming increasingly important

► **NATURE.COM**

For more on university museums:
go.nature.com/mu92rw

GIAMBATTISTA VOLPI (1752–1821) / MUS. NAT. HIST., UNIV. PAVIA

elements of museum management, whether local, regional or national — particularly in the current climate of reduced funding. In Italy, these policies are crucial. The country harbours 47 World Heritage sites, more than any other nation, and is home to 4,000 publicly registered museums, including some of the most important art collections in the world. But funds are shrinking, to the point at which the existence of some museums, and some of the country's most famous works, hangs in the balance. For example, in June it was discovered that Raphael's masterpiece *The Marriage of the Virgin* was endangered by water seeping through the walls of the Brera Art Gallery in Milan, owing to a lack of routine maintenance.

Yet strategic disposals, including long-term or even indefinite loans of museum pieces, are not being seriously considered by museums worldwide as a way to cope with the costs of maintaining full public collections. Museum authorities in Italy, as in many other countries, have elected to adopt such measures as reduced visiting hours rather than consider the disposal of selected holdings publicly and openly — perhaps fearing a public backlash. This is not, of course, to suggest that museums should sell holdings simply to cover the costs of routine maintenance; but it may be necessary for some to be realistic and open about the size of their holdings and the number of pieces that can be successfully exhibited.

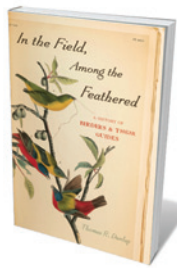
To achieve a balanced ebb and flow of pieces in museum collections, the phases of accession and disposal need to be carefully evaluated and managed. Traditionally, curators tended not to limit the inflow of items — in some cases coming to resemble compulsive hoarders. Today it is common for curators, applying narrower and more demanding criteria in their selection of accessions, to refuse 90% of the offers made.

Although most donations may be turned down without damaging the public perception of a museum, disposal of materials that are already held is a different matter. In the United Kingdom in 2009–10, for instance, public outcry halted Southampton City Council's attempt to sell works by sculptor Auguste Rodin and British artist Alfred Munnings in order to raise money for a new museum dedicated to the RMS *Titanic*. Competing resources, and debates over the relative significance of pieces or collections, form just part of the criteria for deaccession. The views of museum audiences need to be factored in.

Debates over deaccession and disposal should be conducted transparently, accompanied by a clear public explanation and rationale, and should involve the views of all stakeholders — from museum staff to local authorities and concerned individuals. This way, the public's sense of ownership of prized local collections can be honoured.

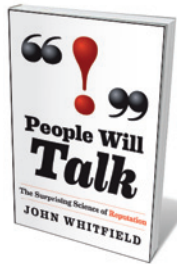
In the 1980s and early 1990s, Canada's ►

Books in brief



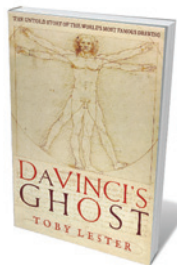
In the Field, Among the Feathered: A History of Birders and Their Guides

Thomas R. Dunlap OXFORD UNIVERSITY PRESS 256 pp. \$34.95 (2011)
Birdwatching may be an amateur science, but it retains close links to ornithology and conservation. In the century since twitching emerged as a suburban hobby, the one constant has been the field guide. Historian and birder Thomas Dunlap trains his binoculars on its evolution in the United States. His meticulous chronicle reveals how the collective urge to spot Nashville warblers or bald eagles ultimately jump-started the US eco-movement, and why the hobby is one of the great unfolding stories of crowd-sourced nature study.



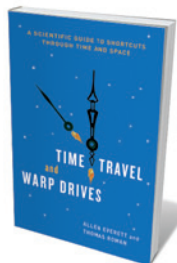
People Will Talk: The Surprising Science of Reputation

John Whitfield WILEY 272 pp. £17.99 (2011)
Reputation is central to our lives, says science writer John Whitfield, yet we are not entirely aware of it. He draws on philosophy, sociology, economics and animal-behaviour experiments to punch through the surface of this powerful and perplexing phenomenon. At base, he finds, it rests on how we trust others and how we persuade others to trust us. An individual's concern for his or her reputation can be a fulcrum that tips them towards either brutal or altruistic behaviour — so ultimately, says Whitfield, the way in which society uses reputation determines whether that society thrives or fails.



Da Vinci's Ghost: The Untold Story of the World's Most Famous Drawing

Toby Lester PROFILE/FREE PRESS 263 pp. £16.99/\$26.99 (2011/2012)
Leonardo da Vinci's *Vitruvian Man* — poised within a circle and a square — is arguably history's most iconic image. Writer Toby Lester offers the absorbing story of this Renaissance rendering. Touching on anatomy, medicine, geography, mathematics, philosophy and aesthetics, he explores the idea that the body, geometry and mystic reality are linked. Its progenitor was Roman architect Marcus Vitruvius Pollio, who posited that human proportions echo the cosmos and should set the form for architecture and for all civilization.



Time Travel and Warp Drives: A Scientific Guide to Shortcuts through Time and Space

Allen Everett and Thomas Roman UNIVERSITY OF CHICAGO PRESS 280 pp. \$30 (2011)
Physicist Allen Everett and mathematician Thomas Roman boldly go into the big speculations that lie at the root of 'frontier science', such as time travel. Through clear explanations and judiciously deployed basic algebra, they lay out the science behind the weirder possibilities offered by the characteristics of space-time. Travelling backwards in time, for instance, can be linked theoretically to moving faster than the speed of light. Warp bubbles, anyone?



Panda: Back from the Brink

Zhou Mengqi SARABAND 192 pp. £25 (2011)
The one-note diet, monochrome markings and long association with fertility treatments in captivity lend the giant panda an unreal air. But these animals are resilient: against all the odds, they total an estimated 2,000 individuals. This tribute to the beauty of the beast is informed by leading conservationists, but centres on stunning shots by photographer Zhou Mengqi. Backed by advice from panda experts Zhang Zhihe, Zhang Hemin and Hu Jinchu, Mengqi tracked his elusive subjects through the mountains of Sichuan for years.

► Glenbow Museum in Calgary, Alberta, successfully used such a process to transfer or sell around 30,000 inherited objects that were no longer relevant to its mission. This large-scale, strategic disposal of objects, including texts and ethnological and military-history collections, allowed the museum to sharpen its focus and release resources for other core sections. Importantly, the move was recognized as valid by the public. The distribution of pieces also boosted the value of the collections they joined.

Deaccession is never simple, and is not to be undertaken lightly. Hastily conceived action could mean big losses for future generations. But meticulously planned disposals for the right reasons are preferable to the disposal of an entire collection, as has happened in the past.

Happily, such disposals may not be the final curtain in every case. Spallanzani's great museum in Pavia may have a future. The collection — a once-cherished jewel of our city and country that has lain dormant for more

than 70 years — was only disbanded, not permanently disposed of. We, at the University of Pavia, are about to begin the process of reconstituting a great part of the original collection for eventual display here, to grant a future to a vital window onto the past. ■

Paolo Mazzarello is director of the University History Museum and head of the University Museums System, University of Pavia, Italy.
e-mail: paolo.mazzarello@unipv.it

CHEMISTRY

An elemental heroine

An opera on the astonishing life of Marie Curie enthralls **Stefan Michalowski** and **Georgia Smith**.

Marie Curie's life seems to be ready-made for grand opera: the brilliant heroine who overcomes poverty and prejudice to marry her soulmate and work with him as an equal, winning a Nobel prize. Then, the tragedy of her husband's accidental death, the scandal of her affair with a married colleague, a second Nobel, heroic war work and her own sad death caused by years of exposure to radiation. Now it is indeed an opera: Polish composer Elżbieta Sikora's *Madame Curie* premiered in Paris on 15 November, and was performed in Gdansk, Poland, on 23–26 November.

Director Marek Weiss brought a full orchestra, cast and chorus from the Baltic Opera of Gdansk to the headquarters of the United Nations Educational, Scientific and Cultural Organization (UNESCO) in Paris for the opening, as part of the International Year of Chemistry.

Sikora's work has been widely performed in contemporary-music venues. She has written large choral pieces, but her two previous operas are on a smaller scale than *Madame Curie*, which has a big, ambitious score that is as precisely wrought as a chamber work. The music is atonal, but is full of attractive melodic structures.

Intriguing orchestral textures are seamlessly integrated with pre-recorded electro-acoustic segments, which Sikora created in collaboration with Argentinian composer Diego Losa. Conductor Wojciech Michniewski elicited a fine performance from the orchestra and chorus in Paris. The cast of fresh-voiced young singers acquitted themselves well, particularly soprano Anna Mikołajczyk as Marie.

The opera does not delve into Curie's scientific discoveries, nor does it try to depict



Soprano Anna Mikołajczyk as Marie Curie.

the excitement that gripped the research community during the birth of subatomic physics in the early 1900s. Sikora said in a pre-show interview that she “was looking for a strong woman character”, and although the stage set is the Curies' rudimentary laboratory, it is Marie as a heroic woman that we see, rather than Marie as a brilliant scientific mind.

As the opera opens, she enters in her nightgown, reading a letter from the Swedish Academy that asks her not to come to Stockholm to accept her second Nobel prize because of the public scandal about her affair with physicist Paul Langevin. The rest of the opera is a succession of scenes from her life,

presented with little or no visual demarcation. Weiss explained in an interview that he envisioned the opera as a dream that Marie has after receiving the letter. This didn't

come across clearly on stage, however, and at first the plot was confusing. But once the spectator gets used to the impressionist-fantasy narrative style, it works.

Weiss has created several beautifully effective scenes, notably Marie's grief after Pierre's death, a comic duel over the Langevin affair, and scenes in which the chorus plays a fickle public vacillating between idolizing Curie and despising her. The finale is particularly powerful: as the orchestra builds a stunning crescendo, Marie hobbles from the stage supporting the lifeless body of a dancer who has been her symbolic double throughout.

The music for the voices spans the spectrum from intense lyricism to plain speech. Marie's aria lamenting Pierre's death, movingly performed by Mikołajczyk, is a striking combination of late-Romantic aesthetic articulated through the idiom of contemporary music. Sikora has a flair for the sudden quiet of instrumental solos: one dance sequence is accompanied only by a long, ravishing clarinet solo. More such moments might be welcome in what is otherwise a relentless pattern of repeated, hard-hitting climaxes in both drama and music.

There is one serious disappointment in the way this story is told. Directly after the overture, an aged man resembling Albert Einstein appears to Marie. He warns her that devastating consequences could result from her work, and urges her to stop. A video projection of a mushroom cloud rams home the point.

Marie shakes off this vision, insisting on the necessity of pursuing the truth. But associating Curie with nuclear weapons is untenable — the relevant discoveries were made shortly before and after her death. Attributing the ethical choices of later scientists to her makes no more sense than the dismissive sexism she endured during her lifetime. ■

Stefan Michalowski is a former particle physicist, and executive secretary of the OECD Global Science Forum in Paris.

Georgia Smith is a freelance journalist based in Paris.
e-mails: stefanm@noos.fr; georgias@noos.fr

This article does not represent the views of the OECD.

S. CŹWIKŁA/OPERA BALTYCKA

elements of museum management, whether local, regional or national — particularly in the current climate of reduced funding. In Italy, these policies are crucial. The country harbours 47 World Heritage sites, more than any other nation, and is home to 4,000 publicly registered museums, including some of the most important art collections in the world. But funds are shrinking, to the point at which the existence of some museums, and some of the country's most famous works, hangs in the balance. For example, in June it was discovered that Raphael's masterpiece *The Marriage of the Virgin* was endangered by water seeping through the walls of the Brera Art Gallery in Milan, owing to a lack of routine maintenance.

Yet strategic disposals, including long-term or even indefinite loans of museum pieces, are not being seriously considered by museums worldwide as a way to cope with the costs of maintaining full public collections. Museum authorities in Italy, as in many other countries, have elected to adopt such measures as reduced visiting hours rather than consider the disposal of selected holdings publicly and openly — perhaps fearing a public backlash. This is not, of course, to suggest that museums should sell holdings simply to cover the costs of routine maintenance; but it may be necessary for some to be realistic and open about the size of their holdings and the number of pieces that can be successfully exhibited.

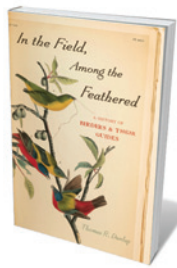
To achieve a balanced ebb and flow of pieces in museum collections, the phases of accession and disposal need to be carefully evaluated and managed. Traditionally, curators tended not to limit the inflow of items — in some cases coming to resemble compulsive hoarders. Today it is common for curators, applying narrower and more demanding criteria in their selection of accessions, to refuse 90% of the offers made.

Although most donations may be turned down without damaging the public perception of a museum, disposal of materials that are already held is a different matter. In the United Kingdom in 2009–10, for instance, public outcry halted Southampton City Council's attempt to sell works by sculptor Auguste Rodin and British artist Alfred Munnings in order to raise money for a new museum dedicated to the RMS *Titanic*. Competing resources, and debates over the relative significance of pieces or collections, form just part of the criteria for deaccession. The views of museum audiences need to be factored in.

Debates over deaccession and disposal should be conducted transparently, accompanied by a clear public explanation and rationale, and should involve the views of all stakeholders — from museum staff to local authorities and concerned individuals. This way, the public's sense of ownership of prized local collections can be honoured.

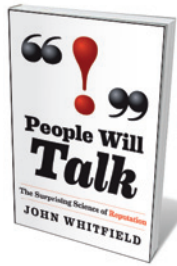
In the 1980s and early 1990s, Canada's ►

Books in brief



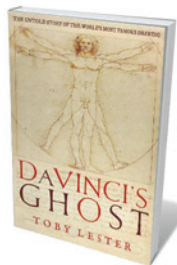
In the Field, Among the Feathered: A History of Birders and Their Guides

Thomas R. Dunlap OXFORD UNIVERSITY PRESS 256 pp. \$34.95 (2011)
Birdwatching may be an amateur science, but it retains close links to ornithology and conservation. In the century since twitching emerged as a suburban hobby, the one constant has been the field guide. Historian and birder Thomas Dunlap trains his binoculars on its evolution in the United States. His meticulous chronicle reveals how the collective urge to spot Nashville warblers or bald eagles ultimately jump-started the US eco-movement, and why the hobby is one of the great unfolding stories of crowd-sourced nature study.



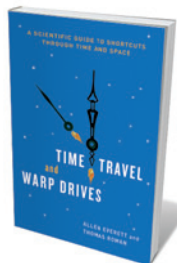
People Will Talk: The Surprising Science of Reputation

John Whitfield WILEY 272 pp. £17.99 (2011)
Reputation is central to our lives, says science writer John Whitfield, yet we are not entirely aware of it. He draws on philosophy, sociology, economics and animal-behaviour experiments to punch through the surface of this powerful and perplexing phenomenon. At base, he finds, it rests on how we trust others and how we persuade others to trust us. An individual's concern for his or her reputation can be a fulcrum that tips them towards either brutal or altruistic behaviour — so ultimately, says Whitfield, the way in which society uses reputation determines whether that society thrives or fails.



Da Vinci's Ghost: The Untold Story of the World's Most Famous Drawing

Toby Lester PROFILE/FREE PRESS 263 pp. £16.99/\$26.99 (2011/2012)
Leonardo da Vinci's *Vitruvian Man* — poised within a circle and a square — is arguably history's most iconic image. Writer Toby Lester offers the absorbing story of this Renaissance rendering. Touching on anatomy, medicine, geography, mathematics, philosophy and aesthetics, he explores the idea that the body, geometry and mystic reality are linked. Its progenitor was Roman architect Marcus Vitruvius Pollio, who posited that human proportions echo the cosmos and should set the form for architecture and for all civilization.



Time Travel and Warp Drives: A Scientific Guide to Shortcuts through Time and Space

Allen Everett and Thomas Roman UNIVERSITY OF CHICAGO PRESS 280 pp. \$30 (2011)
Physicist Allen Everett and mathematician Thomas Roman boldly go into the big speculations that lie at the root of 'frontier science', such as time travel. Through clear explanations and judiciously deployed basic algebra, they lay out the science behind the weirder possibilities offered by the characteristics of space-time. Travelling backwards in time, for instance, can be linked theoretically to moving faster than the speed of light. Warp bubbles, anyone?



Panda: Back from the Brink

Zhou Mengqi SARABAND 192 pp. £25 (2011)
The one-note diet, monochrome markings and long association with fertility treatments in captivity lend the giant panda an unreal air. But these animals are resilient: against all the odds, they total an estimated 2,000 individuals. This tribute to the beauty of the beast is informed by leading conservationists, but centres on stunning shots by photographer Zhou Mengqi. Backed by advice from panda experts Zhang Zhihe, Zhang Hemin and Hu Jinchu, Mengqi tracked his elusive subjects through the mountains of Sichuan for years.

► Glenbow Museum in Calgary, Alberta, successfully used such a process to transfer or sell around 30,000 inherited objects that were no longer relevant to its mission. This large-scale, strategic disposal of objects, including texts and ethnological and military-history collections, allowed the museum to sharpen its focus and release resources for other core sections. Importantly, the move was recognized as valid by the public. The distribution of pieces also boosted the value of the collections they joined.

Deaccession is never simple, and is not to be undertaken lightly. Hastily conceived action could mean big losses for future generations. But meticulously planned disposals for the right reasons are preferable to the disposal of an entire collection, as has happened in the past.

Happily, such disposals may not be the final curtain in every case. Spallanzani's great museum in Pavia may have a future. The collection — a once-cherished jewel of our city and country that has lain dormant for more

than 70 years — was only disbanded, not permanently disposed of. We, at the University of Pavia, are about to begin the process of reconstituting a great part of the original collection for eventual display here, to grant a future to a vital window onto the past. ■

Paolo Mazzarello is director of the University History Museum and head of the University Museums System, University of Pavia, Italy.
e-mail: paolo.mazzarello@unipv.it

CHEMISTRY

An elemental heroine

An opera on the astonishing life of Marie Curie enthralls **Stefan Michalowski** and **Georgia Smith**.

Marie Curie's life seems to be ready-made for grand opera: the brilliant heroine who overcomes poverty and prejudice to marry her soulmate and work with him as an equal, winning a Nobel prize. Then, the tragedy of her husband's accidental death, the scandal of her affair with a married colleague, a second Nobel, heroic war work and her own sad death caused by years of exposure to radiation. Now it is indeed an opera: Polish composer Elżbieta Sikora's *Madame Curie* premiered in Paris on 15 November, and was performed in Gdansk, Poland, on 23–26 November.

Director Marek Weiss brought a full orchestra, cast and chorus from the Baltic Opera of Gdansk to the headquarters of the United Nations Educational, Scientific and Cultural Organization (UNESCO) in Paris for the opening, as part of the International Year of Chemistry.

Sikora's work has been widely performed in contemporary-music venues. She has written large choral pieces, but her two previous operas are on a smaller scale than *Madame Curie*, which has a big, ambitious score that is as precisely wrought as a chamber work. The music is atonal, but is full of attractive melodic structures.

Intriguing orchestral textures are seamlessly integrated with pre-recorded electro-acoustic segments, which Sikora created in collaboration with Argentinian composer Diego Losa. Conductor Wojciech Michniewski elicited a fine performance from the orchestra and chorus in Paris. The cast of fresh-voiced young singers acquitted themselves well, particularly soprano Anna Mikołajczyk as Marie.

The opera does not delve into Curie's scientific discoveries, nor does it try to depict



Soprano Anna Mikołajczyk as Marie Curie.

the excitement that gripped the research community during the birth of subatomic physics in the early 1900s. Sikora said in a pre-show interview that she “was looking for a strong woman character”, and although the stage set is the Curies' rudimentary laboratory, it is Marie as a heroic woman that we see, rather than Marie as a brilliant scientific mind.

As the opera opens, she enters in her nightgown, reading a letter from the Swedish Academy that asks her not to come to Stockholm to accept her second Nobel prize because of the public scandal about her affair with physicist Paul Langevin. The rest of the opera is a succession of scenes from her life,

presented with little or no visual demarcation. Weiss explained in an interview that he envisioned the opera as a dream that Marie has after receiving the letter. This didn't

come across clearly on stage, however, and at first the plot was confusing. But once the spectator gets used to the impressionist-fantasy narrative style, it works.

Weiss has created several beautifully effective scenes, notably Marie's grief after Pierre's death, a comic duel over the Langevin affair, and scenes in which the chorus plays a fickle public vacillating between idolizing Curie and despising her. The finale is particularly powerful: as the orchestra builds a stunning crescendo, Marie hobbles from the stage supporting the lifeless body of a dancer who has been her symbolic double throughout.

The music for the voices spans the spectrum from intense lyricism to plain speech. Marie's aria lamenting Pierre's death, movingly performed by Mikołajczyk, is a striking combination of late-Romantic aesthetic articulated through the idiom of contemporary music. Sikora has a flair for the sudden quiet of instrumental solos: one dance sequence is accompanied only by a long, ravishing clarinet solo. More such moments might be welcome in what is otherwise a relentless pattern of repeated, hard-hitting climaxes in both drama and music.

There is one serious disappointment in the way this story is told. Directly after the overture, an aged man resembling Albert Einstein appears to Marie. He warns her that devastating consequences could result from her work, and urges her to stop. A video projection of a mushroom cloud rams home the point.

Marie shakes off this vision, insisting on the necessity of pursuing the truth. But associating Curie with nuclear weapons is untenable — the relevant discoveries were made shortly before and after her death. Attributing the ethical choices of later scientists to her makes no more sense than the dismissive sexism she endured during her lifetime. ■

Stefan Michalowski is a former particle physicist, and executive secretary of the OECD Global Science Forum in Paris.

Georgia Smith is a freelance journalist based in Paris.
e-mails: stefanm@noos.fr; georgias@noos.fr

This article does not represent the views of the OECD.

S. CÍWIKŁA/OPERA BALTYCKA

Correspondence

Arab science to turn spring into summer

For Arab scientists, the social revolution of the past year could be the light at the end of a long tunnel of hopelessness. To transform this optimistic mood into action, we must consider what we need to do to integrate science and technology into post-revolution development models, and what we can realistically accomplish.

The issue of governance is high on the agenda: it should be rooted in democracy and merit from now on. Innovative national systems are needed that reject trade-offs between educational reform and investment in education, research and design.

It is important that new science and technology development models are adapted to the socioeconomic environment and resist the temptation to enter into blind competition with developed countries.

Because scientific expertise will be needed to inspire these plans, it is imperative that models include strategies to reverse the scientific diaspora and to prevent the brain drain from Arab countries.

M. Dahmani Fathallah *Arabian Gulf University, Bahrain; and Arab Policy Institute, Tunisia.*
d.fathallah@agu.edu.bh

Climate results for public vetting

Your discussion of the media's interest in the Berkeley Earth Surface Temperature (BEST) study — which released its results before peer review — implicitly acknowledges a change in the public authorization of scientific knowledge (*Nature* **478**, 428; 2011). Yet few scientists would accept your branding of them as “purist” or “renegade” for their stance on the study's approach.

BEST members seek to promote a progressive image of science in a demanding, open and transparent society. Stimulated by an erosion of public trust

in science, the publication of research before peer review aims to satisfy calls for greater public engagement with science.

Peer review after publication won't just confirm, adjust or correct the BEST results: it will demarcate soberly vetted scientific knowledge from the sometimes glib inferences an emotionally charged public makes through ‘unruly’ media that do not always report as scientists might want them to.

The progressive publication strategy of the BEST study exemplifies the emergence of a double-truth approach to the authorization of scientific knowledge on climate change.
Mathis Hampel, Martin Mahony, Mike Hulme *University of East Anglia, Norwich, UK.*
mathis.hampel@gmail.com

Control the bonanza for research eggs

The demand for women's eggs for research could soar alarmingly following news of a cloning technique that uses human oocytes to reprogram somatic cells to a state of pluripotency (S. Noggle *et al.* *Nature* **478**, 70–75; 2011).

The mean number of eggs given by each woman during the study was 16.9, with one donating 26 eggs. This is more than many fertility doctors would consider optimal and increases the risk of ovarian hyperstimulation syndrome. The researchers do not say that they halted hormone treatment in cases of over-response, although they did stop it in under-responsive women.

Noggle *et al.* rightly anticipated concerns that payment for eggs could encourage financially disadvantaged women to take risks they might otherwise avoid. But US\$8,000, the amount paid by Noggle and colleagues, would be a temptation even to the well-off in these difficult economic times.

Some argue that women should evaluate for themselves the risks and benefits of providing eggs for

research. But informed consent depends on provision of accurate information. Even after years of egg harvesting for fertility treatment, the risks to women — especially from some of the drugs and hormones used — remain undercharacterized and poorly assessed, with inadequate follow-up and data collection.

Marcy Darnovsky *Center for Genetics and Society, Berkeley, California, USA.*
mdarnovsky@geneticsandsociety.org
Susan Berke Fogel *Pro-Choice Alliance for Responsible Research, Van Nuys, California, USA.*

Judy Norsigian *Our Bodies Ourselves, Cambridge, Massachusetts, USA.*

Food security: a role for Europe

The European Union (EU) could contribute considerably to sustainable global food security because of its wealth of productive farmland (J. A. Foley *et al.* *Nature* **478**, 337–342; 2011). Both production growth and increased sustainability therefore need to be at the heart of the Common Agricultural Policy.

As one of the largest global importers and exporters of food, the EU has significant leverage in, and responsibility for, the global food and nutrition agenda. Although the biggest impact on global food security will result from increased production in developing countries, the EU must play its part. It needs to protect and maintain its agricultural lands and increase food production.

This will mean producing more food with fewer resources, while minimizing environmental impacts at multiple spatial and temporal scales (called ‘sustainable intensification’). Biotechnology will have a role in global food security, but it is unlikely to be the main solution. New approaches to managing farming systems and agricultural landscapes in an environmentally sensitive way could instead be more productive.

We therefore welcome the European Commission's proposals to increase the budget for agricultural research and innovation that is targeted at productive and ecological systems.

Tim Benton *University of Leeds, UK.*
t.g.benton@leeds.ac.uk
Tibor Hartel *Ovidius University of Constanta, Romania.*

Josef Settele *Helmholtz Centre for Environmental Research — UFZ, Halle, Germany.*
On behalf of the other 6 authors of the Warsaw Consensus Statement (see go.nature.com/jzuytm).

Food security: close crop yield gap

Crop yields are rising more slowly than they were several decades ago (J. A. Foley *et al.* *Nature* **478**, 337–342; 2011). This is especially evident in Europe, and has serious implications for its responsibilities in global food production.

In Denmark, France, Finland and Switzerland (see, for example, R. Finger *Food Policy* **35**, 175–182; 2010), the ‘yield gap’ has increased: the rate of growth in crop yields has declined, even though yield potential has risen over the past few decades owing to technological advances such as crop breeding.

Markets are one contributory factor: lower prices caused by reduced market support have decreased the incentive for investment in equipment, fertilizer, and so on. Also, by restricting the use of pesticides, for example, agricultural policies aimed at reducing environmental damage have hindered growth in crop yields.

Closing these yield gaps will ensure sufficient global food production and help towards food security. More incentives are needed, particularly as low-income countries are also confronted by sizeable yield gaps, for different reasons.

Robert Finger *ETH Zurich, Switzerland.*
rofinger@ethz.ch

John McCarthy

(1927–2011)

The scientist who set computers on the path to common sense.

In 1955, a group of scientists approached the Rockefeller Foundation with a request to fund “a 2 month, 10 man study of artificial intelligence” to be carried out at Dartmouth College in New Hampshire. This was the earliest use of the term ‘artificial intelligence’, which was coined by one of the proposers — John McCarthy, then an assistant professor of mathematics at Dartmouth.

McCarthy was born in 1927 in Boston, Massachusetts, to an Irish immigrant father and a Lithuanian Jewish immigrant mother. When he was accepted into the California Institute of Technology in Pasadena at the age of 16, McCarthy was able to skip the first two years of mathematics because he had taught himself calculus from textbooks. He received a PhD in mathematics from Princeton University and, after a few other short appointments, became a full professor at Stanford University in California in 1962.

Computer scientists will remember McCarthy as the creator of LISP, one of the oldest programming languages still in wide use today, and as the inventor of time-sharing technology — the sharing of a computing resource among many users.

But his most notable contributions were in artificial intelligence. McCarthy’s long-term goal was to design a computer program that could derive logical consequences from anything it was told, in combination with its previous knowledge, and act on the basis of its conclusions — a property, he observed, that is similar to “what makes us describe certain humans as having common sense”. This research programme was described in his 1959 symposium paper ‘Programs with common sense’, often referred to as the ‘advice-taker’ paper, after McCarthy’s name for the proposed computer program.

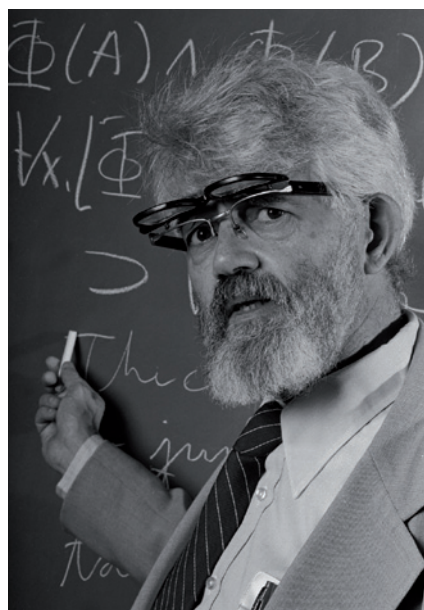
The advice-taker paper quickly became one of the most influential publications in the field. The origins of several branches of artificial intelligence can be traced back to it.

“McCarthy, the consummate independent thinker, never shied away from speaking his mind.”

One of these branches is ‘automated planning’, which aims to generate a strategy that would allow an agent, such as a robot or an unmanned vehicle, to achieve a given goal. In the advice-taker paper, McCarthy gives the example of being seated at a desk at home, where one

has a car, and wishing to go to the airport. His common-sense computer program, he posited, should be able to find a simple two-step solution: walk to the car and drive to the airport. This was the first published example of an automated planning problem.

Another sub-area that finds its roots in McCarthy’s early work is that of ‘knowledge representation’: the concept of coding knowledge in formal notation so that computer programs can easily make inferences from



it. McCarthy’s program, for example, would need some knowledge about the distance from home to the airport, to infer that this distance is too far to walk. This differs from most elements of a traditional computer program, which are imperative statements such as: ‘increase the value of a variable, N , by one’. The advice-taker paper instead stresses the role of declarative sentences.

One of McCarthy’s breakthroughs was his theory of circumscription: a mathematical theory of exceptions to general statements. McCarthy realized that the general assertions we make in everyday life are true only in typical cases, and allow for occasional exceptions: as in the statement “I always eat lunch at noon”, for example. In mathematics, however, ‘always’ means literally no exceptions. McCarthy’s theory allowed a computer program to presume that things are as expected in the absence of information to the contrary, and to retract its conclusions

when it learns about exceptions. This is known as non-monotonic reasoning.

My collaboration with McCarthy began in 1984, when I learned about his work on circumscription and was fascinated by its beauty and intellectual depth. Like dozens of other scientists, I often spent long hours thinking about research problems that he posed. These always touched on fundamental questions of intelligence and often used interesting examples. One familiar logician’s example involves a group of missionaries and cannibals aiming to cross a river, given only a small boat and the problem that the cannibals might eat the missionaries if the latter are outnumbered. McCarthy wondered how a computer should deal with variations on the traditional formulation: what if one of the missionaries is Jesus Christ, who can cross the river without a boat? What if three missionaries alone with a cannibal can convert him into a missionary?

Progress towards programs with common sense was at first painfully slow. Efficient implementations of non-monotonic reasoning only became available about ten years ago and are applied in many areas, including the control of spacecraft. But the full realization of McCarthy’s dream is still far in the future.

McCarthy’s parents, whom he never ceased to love and admire, were members of the Communist Party, and in 1949 he too joined the party. But he broke away from these doctrines a few years later after he subjected them to rational analysis, and became a conservative. The sustainability of human material progress and the desirability of nuclear power were ideas particularly dear to his heart. He opposed affirmative-action policies that led to reverse discrimination, and did not hide his disdain for the sort of political correctness that uses the pronoun ‘she’ when the gender of a person is irrelevant.

In academia, liberals far outnumber conservatives, and it is understandable that many conservative professors prefer to keep their political views to themselves. McCarthy, the consummate independent thinker, never shied away from speaking his mind, both in science and in politics. ■

Vladimir Lifschitz is professor of computer science at the University of Texas at Austin, Texas 78712, USA.
e-mail: vl@cs.utexas.edu

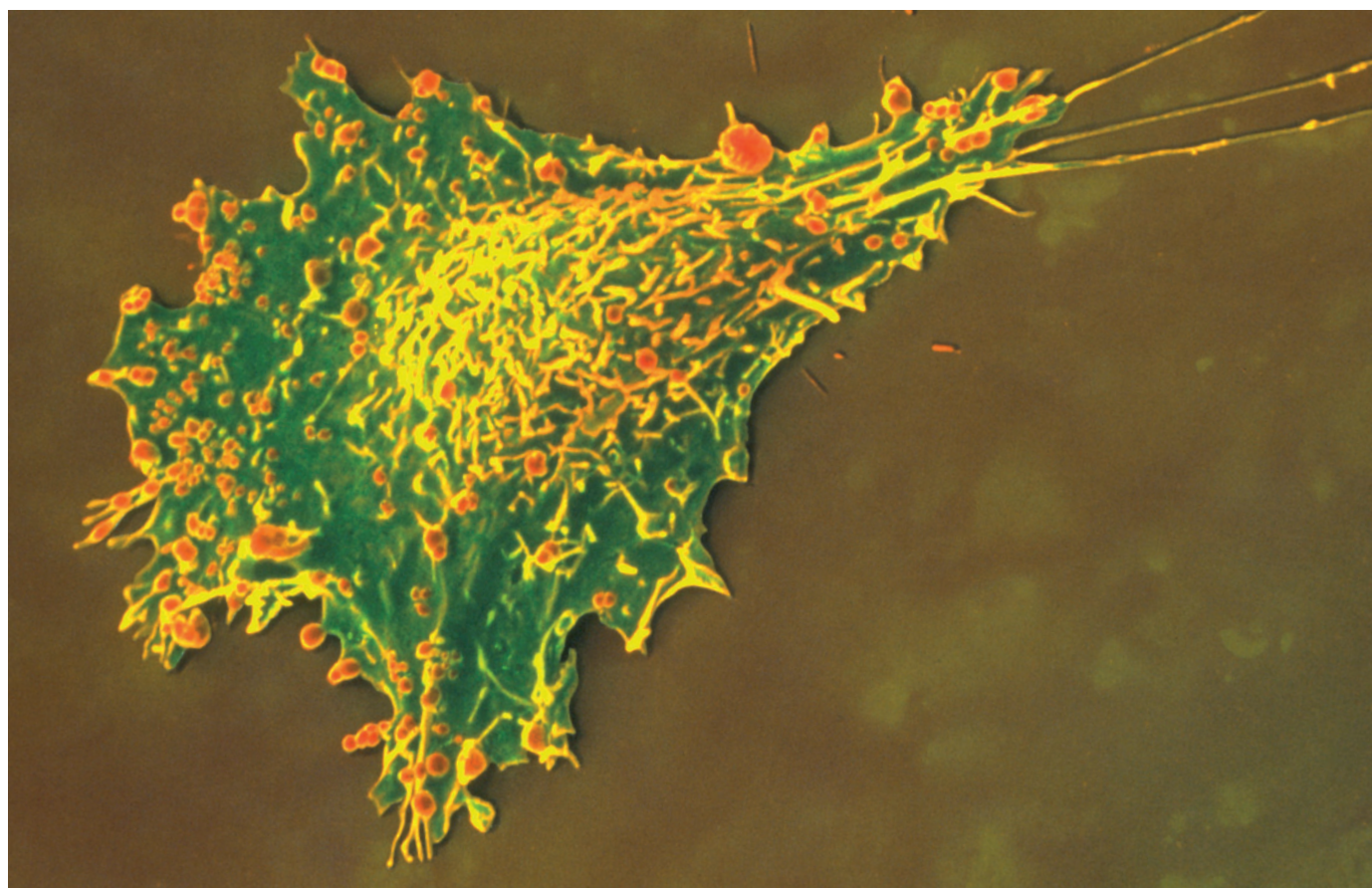
R. RESSMEYER/CORBIS

TECHNOLOGY FEATURE

THE DEEPEST DIFFERENCES

To understand biological heterogeneity, researchers are learning how to profile the molecular contents of individual cells.

S. OSCHMEISSNER/SPL



Studies are uncovering the molecular biology of individual tumour cells, such as this breast-cancer cell, seen in a coloured scanning electron micrograph.

BY CHARLOTTE SCHUBERT

James Eberwine, a neuroscientist with a penchant for invention, helped to pioneer a technique that is now routine. In the early 1990s, he sucked the contents out of a single cell with a pipette, and examined the expression of a handful of genes using molecular techniques that amplify RNA. His data verified a long-held assumption — that electrical activity in a neuron simultaneously changes the abundance of multiple RNAs inside it¹. But other researchers were sceptical. At the time, just about the only way to detect RNA in

a single cell was by labelling molecules using fluorescence *in situ* hybridization. “People were used to using microscopy to look at RNA; they wanted to see it,” says Eberwine, who works at the University of Pennsylvania in Philadelphia.

Things have changed since then. Gene-expression analyses leaped forward in the mid-1990s with the invention of the microarray. And the rise of high-throughput RNA sequencing, or RNA-seq, which spits out the sequences of thousands of cellular RNAs at once, has enabled researchers to reveal the collection of active genes in a cell in a single readout.

Eberwine and others are using sequencing, and techniques such as microfluidics and flow cytometry, to profile single cells — cataloguing RNA molecules, sequencing DNA and even profiling metabolites and peptides.

Studies indicate how strongly cells can show their individuality. Brain cells may express as few as 65% of the same genes as their neighbours, according to an unpublished analysis by Eberwine. In the immune system, cells placed in the same category on the basis of surface markers can express different sets of genes, and have different responses to vaccines². And as tumour cells evolve, their genomes quickly ►

► become twisted in unusual ways.

Single-cell techniques let researchers track and catalogue this heterogeneity. They may be the only way to get at some fundamental questions, such as what makes individual cells different biochemically and functionally. How much is each cell influenced by its microenvironment, and what is the role of stochasticity — random ‘noise’ in the behaviour of cellular molecules?

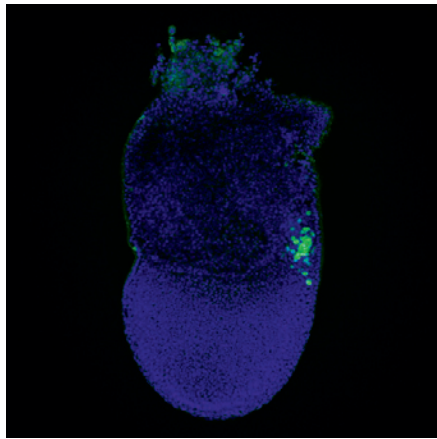
These questions are getting more attention (see page 139). In 2009, Eberwine co-organized a meeting on single-cell analysis at Cold Spring Harbor Laboratory in New York, along with Sunney Xie, a single-cell biochemist at Harvard University in Cambridge, Massachusetts. The meeting drew 47 attendees. This July, 120 people went to the second such meeting. And the US National Institutes of Health (NIH) has launched an initiative to support single-cell techniques (see ‘The NIH gets singular’).

But single-cell analysis is still an emerging field. Many researchers say that protocols from academic labs are often superior to commercial kits. “With any nascent field, there are lots of different approaches,” says Eberwine. “People are trying lots of things to see if they can make the techniques more sensitive, more representative of the state of a cell, easier and cheaper.”

PROFILING HETEROGENEITY

The classic biochemical approach is limited, say single-cell researchers. Grinding up and analysing the contents of large pools of cells — a procedure undertaken in thousands of labs every day — averages out the results, says Timm Schroeder, director of the Institute of Stem Cell Research at the German Center for Environmental Health in Munich. But, says Schroeder, “it’s the individual cell that makes a decision” such as whether to fire an electrical impulse, migrate or differentiate into a new cell type.

Looking at single cells to uncover the impetus for such decisions means doing fussy experiments on a very small object: a cell might span about 10 micrometres and contain less than 1



Studies can probe single cells, such as those in the mouse embryo that give rise to gametes (green).

picolitre (1×10^{-12} litres) of cytoplasm. And some key regulatory molecules are scarce — just a few, hard-to-detect RNAs can exert a big effect on a cell.

Many established techniques are only now being applied to single cells. Fluorescent tagging and microscopy can be used to analyse molecules that have already been characterized. To profile previously unexamined molecules, there is transcriptome analysis — cataloguing the set of RNAs expressed in a cell — as well as high-throughput methods based on microfluidics or flow cytometry. But getting such techniques to work on single cells is not easy, says Schroeder, whose research involves long-term imaging of individual bone-marrow cells. Single-cell applications are, he says, “at least one level more demanding and complex than the conventional approaches”.

And the unexplored biological terrain is vast. “We don’t even know what we are getting into in terms of heterogeneity,” says Sherman Weissman, a geneticist at Yale University in New Haven, Connecticut.

Whatever the study, the first step is generally getting hold of the cells. When Eberwine first struggled with studying gene expression in a neuron 20 years ago, it was difficult even to get

intact RNA out of a single cell. Eberwine solved the problem by capturing the material in the same pipette used to measure electrical activity. Now, researchers can use a variety of techniques to pick out single cells, from enzymatic digestion, which releases cells from tissues, to laser-capture microdissection. But it is still tricky, says Eberwine. “A major technical issue is how do you do that initial capture.”

Eberwine has used his pipette-capture system to study individual warm-sensitive neurons³, which regulate core body temperature and underlie fever. Together with Tamas Bartfai, a neuroscientist at the Scripps Research Institute in La Jolla, California, and his colleagues, Eberwine examined the cells’ transcriptomes. The researchers identified transcripts for G-protein-coupled receptors — potential drug targets — that went undetected in screens of pooled cells.

As techniques improve, they are letting researchers explore the heterogeneity within a cell. By cutting branches, or dendrites, off

“It’s the individual cell that makes a decision.”

neurons, Eberwine and his colleagues have discovered that RNA in dendrites can retain nucleotide sequences that target the RNA to that location⁴. They could not have found such information by analysing the RNA of an entire neuron.

Developing technology will produce even more fresh data. Most biologists will need to work closely with computational biologists to evaluate the huge data sets that will result from cataloguing thousands of molecules in numerous single-cell experiments.

EXTRA-LOUD AMPLIFICATION

Perhaps the best-known single-cell profiling technique is transcriptomics. Azim Surani, a developmental biologist at the University of Cambridge, UK, uses this method to examine cells of the early embryo, which are hard to study in large batches because they are so rare. He is tracing how, such cells turn into pluripotent embryonic stem cells in culture.

Surani has adapted a single-cell protocol for the polymerase chain reaction (PCR) to work with RNA-seq. To do this, he has collaborated with technical experts such as Kaiqin Lao, a molecular cell biologist at Applied Biosystems in Foster City, California (a subsidiary of Life Technologies in Carlsbad). In the cells of the early mouse embryo, the team detected the expression of some 12,300 genes — 75% more than were detected by microarray techniques⁵. Lao says he can now get his PCR technique to work with 1 picogram of RNA — one-tenth the amount of RNA in a typical cell.

The published protocol amplifies molecules only if they are no more than 3 kilobases long, so it misses about 40% of transcripts, says Lao. He and his colleagues are using different

The NIH gets singular

The challenges of single-cell analysis have caught the attention of the US National Institutes of Health (NIH). The agency has launched a programme to fund advances in single-cell research, with a budget of around US\$90 million over five years from the NIH Common Fund, which backs science that crosses disciplines. Grant applications are due early next year, and the NIH expects to make the first awards by September 2012, says Andrea Beckel-Mitchener, a programme officer at the NIH campus in Bethesda,

Maryland. The programme will fund new techniques in areas ranging from microscopy to biochemistry, and foster their commercialization. The NIH also sees a big need for tools to examine cells in their natural environment.

Many of the techniques need an extra push. “It’s still really difficult for individual labs to move into that area; the group of researchers who work on this is still highly specialized,” says Beckel-Mitchener. “If you want to reach the next level you really have to push the envelope.” **C.S.**

enzymes to increase that; Lao can now amplify 10-kilobase transcripts, corresponding to about 99% of transcription, he says.

Another technique to amplify a cell's RNA is antisense RNA (aRNA), an *in vitro* transcription technique from Eberwine and his colleagues⁶, in which a cell's RNA is copied into a stable DNA library, with each DNA molecule containing a short sequence recognized by an RNA polymerase. The polymerase uses the DNA library to make multiple copies of the RNA.

“Every neuron is probably different from every other neuron.”

Each approach has its advantages, and its problems. Bias can be introduced to PCR when certain sequences dominate during amplification, so approaches based on this technique are less quantitative than aRNA. But aRNA is less efficient than PCR, and can take days, notes Weissman.

Commercially available aRNA kits include

TargetAmp from Epicentre Biotechnologies of Madison, Wisconsin (owned by Illumina of San Diego, California), and MessageAmp from Ambion of Austin, Texas, which is owned by Life Technologies. Both can work for single cells, says Eberwine. Companies such as NuGEN in San Carlos, California, and Sigma-Aldrich in St Louis, Missouri (in partnership with Rubicon Genomics of Ann Arbor, Michigan), have products designed for small amounts of RNA, and some say that their systems can work for single cells. It is unclear when Life Technologies might release a product based on Lao's method, but both Eberwine and Lao report that their single-cell techniques are being used successfully in other labs.

THE GENOME GAP

Many researchers want to analyse not just the transcriptome of a cell, but the underlying genome. This would be particularly relevant for cancer cells, with their warped DNA, and Life Technologies is offering US\$1 million to the first researchers to sequence the entire

genome and RNA content of a single cancer cell using the company's technology.

Nicholas Navin, a geneticist at the MD Anderson Cancer Center in Houston, Texas, is one of only a handful of researchers who have sequenced the genomes of single cells from eukaryotic organisms. This year, in collaboration with Michael Wigler, a geneticist at Cold Spring Harbor, and his colleagues, Navin sequenced the DNA of 100 individual cells from different parts of each of two human breast tumours, tracing how the cancer evolves as it spreads⁷. It took several years and cost about \$2,000 per cell; the cost has since fallen to about \$200 per cell, he says. In the end, Navin was able to reliably cover about 6% of the genome of a cell — enough to assess some larger copy-number aberrations, but not to look at the accumulation of point mutations during tumour evolution.

The limitation, say Navin and other researchers, is the technique used to amplify the DNA: whole-genome amplification, which

Beyond amplification

Unlike DNA and RNA, some cellular molecules cannot be amplified. Very few labs have tried to profile proteins, metabolites or peptides in single cells. “The analytical challenges are extreme,” says Renato Zenobi, a chemist at the Swiss Federal Institute of Technology in Zurich.

To assess populations of proteins, some researchers have used fluorescent tagging. Sunney Xie, a single-cell biochemist at Harvard University in Cambridge, Massachusetts, for instance, has harnessed a microfluidic device to quantify the set of proteins and RNA in the *Escherichia coli* bacterium, using a fluorescent-protein reporter library¹².

Sherman Weissman, a geneticist at Yale University in New Haven, Connecticut, has hopes for ribosome profiling, a molecular technique to monitor protein production. The procedure involves obtaining a sequence of RNAs (about 28 nucleotides long) that have been sequestered by protein-making ribosomes^{13,14}. The method works with pools of cells, but it may have potential for single cells, says Weissman.

Daniel Chiu, a chemist at the University of Washington in Seattle, has a device that can add reagents to label a cell or organelle and send its contents — one molecule at a time — to an instrument that can count fluorescent molecules¹¹. His team has hooked this up to another microfluidic system: a high-throughput device to isolate single cells from a population. With this two-step system, the group can isolate cells

and analyse their molecular components. The researchers have adapted the system for the detection and analysis of circulating tumour cells and are launching a company based on the technology: MiCareo, in Taipei, Taiwan.

Researchers working with Garry Nolan, a geneticist at Stanford University in California, and Scott Tanner, a biomedical engineer at the University of Toronto in Canada, are revamping a tried and true high-throughput technique — flow cytometry, which relies on the detection of fluorescent tags, and generally measures six to ten parameters simultaneously. The team has developed ‘mass cytometry’. The technique can simultaneously measure 34 parameters in a single cell¹⁵, including protein phosphorylation and cell-surface antigens, and has the potential for some 100. The instrument that they use evaluates up to 1,000 cells per second, and relies on antibody-bound tags consisting of rare-earth elements detected by mass spectrometry. It is sold by DVS Sciences in Sunnyvale, California, and costs US\$600,000.

Mass spectrometry is routinely used for profiling proteins and metabolites, but does not have the sensitivity of optical techniques such as fluorescence. Very few proteins are present at more than 100,000 copies per cell, so most are below the limit of detection for commercially available mass spectrometers. But researchers are making progress, mainly with more abundant metabolites and peptides. Researchers at the

University of Illinois at Urbana-Champaign, for instance, have used matrix-assisted laser desorption/ionization (MALDI) mass spectrometry to profile the major metabolites and peptide neurotransmitters released in a large neuron of the sea slug *Pleurobranchia californica* in response to hunger¹⁶. In unpublished work, Zenobi and his colleagues have generated a map of the yeast metabolome using MALDI mass-spectrometry data from single cells. He says that he can detect metabolites in the low attomole range (around 1 million molecules), and can identify more than 200 correlations — for instance, whether increases in one metabolite correlate with decreases in another.

Other researchers use different mass spectrometry methods. Cynthia McMurray, a neuroscientist at Lawrence Berkeley National Laboratory in Berkeley, California, is constructing single-cell metabolic maps from thin sections of brain, comparing healthy, aged and diseased mice using nanostructure-initiator mass spectrometry. She ultimately aims to create a three-dimensional metabolic map of the brain.

Zenobi is negotiating with Sigma Aldrich in St Louis, Missouri, about commercializing a micrometre-scale device¹⁷ that prepares multiple single cells on an array for analysis by MALDI mass spectrometry. The device may be available as early as next year. “Once people realize these types of analyses can be done, they will jump on them,” he says. **C.S.**

relies on an enzyme that copies some genomic regions but skips others. By tweaking this step, Navin says, he is now exceeding 50% coverage of the genome of a human cell, although his work has not yet been published.

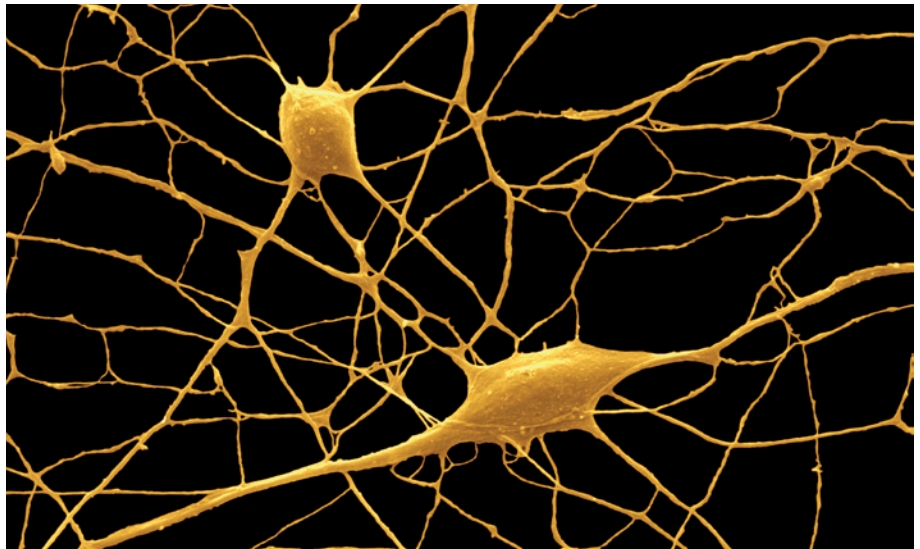
Navin is not the only one tackling this problem. At the Cold Spring Harbor Meeting this year, Xie said that he and his colleagues had been able to sequence 85% of the genome of a mammalian cell. The paper describing it has not yet been published — but researchers who have seen the data are impressed. “Sunney nailed it,” says Lao.

This is welcome news to researchers such as Fred Gage, a neuroscientist at the Salk Institute for Biological Sciences in San Diego, California, who wants to sequence individual neurons. He has found that long interspersed elements (LINEs) — DNA sequences that can move around in the genome — form new insertions when neurons are born from neuronal stem cells⁸. Every neuron probably contains unique LINE insertions, with most cells having between 80 and 300. “Every neuron is probably different from every other neuron,” says Gage.

ONE DEVICE, LOTS OF INFORMATION

Questions on single cells often lead researchers into difficult experimental terrain. For help in navigating such tricky territory, Gage recommends collaborating with the best technical experts; he is working with Roger Lasken, a leader in sequencing unculturable microbes at the J. Craig Venter Institute in San Diego.

But many researchers venturing into single-cell analysis will be on their own, so techniques will have to become more automated, integrated and kit-like, says Jonathan Sweedler, a chemist at the University of Illinois at Urbana–Champaign. “Researchers will be able to buy a device that has 48 steps incorporated into one platform,” he says. Widespread uptake of single-cell analysis will also require high-throughput analyses of dozens to thousands of cells to tease out measurement errors from real heterogeneity.



Researchers are going beyond single cells to profile parts of cells, such as neuronal extensions.

Several companies are working on these goals, offering miniature devices that integrate multiple steps for the high-throughput analysis of single cells. Fluidigm of South San Francisco, California, markets a microfluidic system that can simultaneously analyse 96 genes in 96 individual cells using quantitative PCR. Fluidigm systems have been deployed to uncover previously unrecognized subsets of immune cells², and to examine variability in the response of single cells to cytokine signalling⁹. RainDance Technologies of Lexington, Massachusetts, also sells micro-scale kits to analyse single cells.

Usability and high throughput are a boon for miniature devices. Integrated steps also help to conserve precious samples, and small volumes can aid the dynamics of biochemical reactions — for instance, they can reduce amplification bias in PCR reactions, notes Stephen Quake, a bioengineer at Stanford University in California, and co-founder of Fluidigm.

“Working with small volumes gives you some real technical advantages,” says Quake, whose lab is harnessing microfluidics to develop a technique for single-cell transcriptomics, and has created a device to isolate and sequence single chromosomes¹⁰.

PUTTING IT ALL TOGETHER

But high-throughput techniques will be limited if what they measure is too simple. To grasp how a cell works, “you need to understand not just chemistry, but spatial and temporal information”, says Daniel Chiu, a chemist at the University of Washington in Seattle. To integrate these analyses, his lab combines microfluidics, nanomaterials and optics.

Chiu’s team has developed a technique for single-cell nanosurgery using a ‘vortex trap’, an optical method that can manipulate organelles or liquid droplets. The group has isolated single mitochondria from cells and prepared them for analysis on a ‘droplet nanolab’, which

deploys the vortex trap to fuse droplets and change the concentration of reagents¹¹.

Chiu’s lab has also developed microfluidic devices for quantifying fluorescently tagged molecules, and for detecting and analysing cells that are rare in a population, such as tumour cells circulating in the blood (see ‘Beyond amplification’).

Ultimately, a combination of techniques will be necessary for researchers to attain their goal of measuring multiple parameters in a single, living cell. “The more parameters you can define — the transcriptome, the peptide-ome, how a cell looks, how it responds to drugs — the more information you are going to get out,” says Eberwine.

Eberwine is confident that these methods will emerge, even if it takes years. “I think we will be successful,” he says, “and if we are not, somebody else will be.” ■

Charlotte Schubert is a freelance writer based in Seattle, Washington.

1. Mackler, S. A., Brooks, B. P. & Eberwine, J. H. *Neuron* **9**, 539–548 (1992).
2. Flatz, L. et al. *Proc. Natl Acad. Sci. USA* **108**, 5724–5729 (2011).
3. Eberwine, J. & Bartfai, T. *Pharmacol. Ther.* **129**, 241–259 (2011).
4. Buckley, P. T. et al. *Neuron* **69**, 877–884 (2011).
5. Tang, F. et al. *Nature Meth.* **6**, 377–382 (2009).
6. Van Gelder, R. N. et al. *Proc. Natl Acad. Sci. USA* **87**, 1663–1667 (1990).
7. Navin, N. et al. *Nature* **472**, 90–94 (2011).
8. Coufal, N. G. et al. *Nature* **460**, 1127–1131 (2009).
9. Tay, S. et al. *Nature* **466**, 267–271 (2010).
10. Fan, H. C., Wang, J., Potanina, A. & Quake, S. R. *Nature Biotechnol.* **29**, 51–57 (2011).
11. Chiu, D. T. & Lorenz, R. M. *Acc. Chem. Res.* **42**, 649–658 (2009).
12. Taniguchi, Y. *Science* **329**, 533–538 (2010).
13. Ingolia, N. T., Lareau, L. F. & Weissman, J. S. *Cell* **147**, 789–802 (2011).
14. Ingolia, N. T., Ghaemmaghami, S., Newman, J. R. S. & Weissman, J. S. *Science* **324**, 218–223 (2009).
15. Bendall, S. C. et al. *Science* **332**, 687–696 (2011).
16. Hatcher, N. G. et al. *J. Neurochem.* **104**, 1358–1363 (2008).
17. Urban, P. L. *Lab Chip* **10**, 3206–3209 (2010).



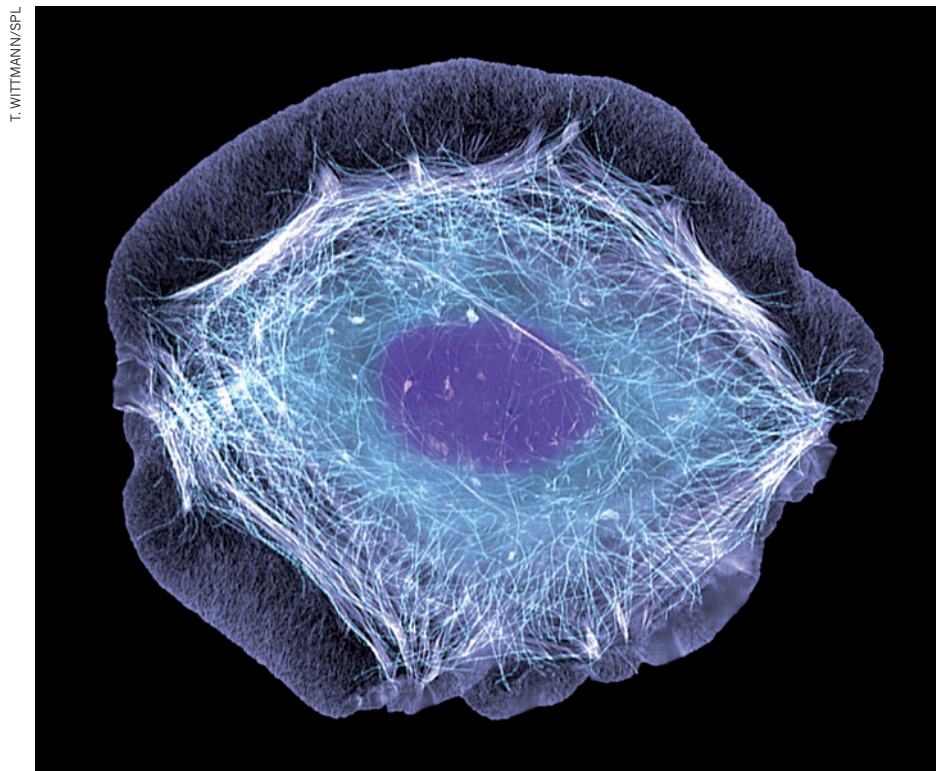
James Eberwine’s ingenuity in pioneering single-cell protocols has led to dozens of patents.

CAREERS

TURNING POINT Former athlete's solar-telescope designs lead to award **p.141**

CAREERS FORUM Get expert advice in an interactive exchange go.nature.com/lm1x4t

NATUREJOBS For the latest career listings and advice www.naturejobs.com



Individual cells, such as this immortal human skin cell, can now be studied in unprecedented detail.

SINGLE-CELL ANALYSIS

Imaging is everything

Advances in single-cell imaging bring opportunities for physicists, biologists and chemists alike.

BY AMY MAXMEN

Enrico Gratton gave his research centre an aspirational and prescient name: the Laboratory for Fluorescence Dynamics. When the lab was established at the University of Illinois at Urbana-Champaign in 1986, Gratton and his colleagues could make lipids and proteins in clumps of cells glimmer using fluorescent probes, but imaging moving molecules in individual cells was impossible. Yet Gratton had hope: “We put ‘dynamics’ in the

centre’s name because we knew we wanted to eventually follow the behaviour of molecules inside individual cells,” he says. With technological advances, Gratton can now do just that at the lab, which moved to the University of California, Irvine, in 2006. He can watch lipids aggregate and proteins exit membranes. His dream has been realized.

Gratton and his colleagues are at the centre of a growing field focused on single cells. Improvements in microscopes, lasers, cameras, computers and fluorescent labels let researchers

directly observe and measure dynamic cellular processes (see page 133). Scientists can look at how tumour cells change over time, measure how long it takes for DNA to bind to a transcription factor or trace the route of RNA as it exits the nucleus, for example.

Since Gratton’s lab became the first single-cell-imaging centre to receive US federal funding, many facilities have opened around the world — including at the German Research Center for Environmental Health in Munich and the Max Planck Institute for Biophysical Chemistry in Göttingen, Germany. “Institutions are quite eager to set up these cores because there’s a huge pay-off,” says Robert Singer, co-director of the Gruss-Lipper Biophotonics Center at the Albert Einstein College of Medicine in New York. “With the right set-up, their faculty can write more-competitive grants and get their studies into better journals.” Investigators can reveal previously unknown biology, and contribute to medical research. A single-cell specialist might reveal, for example, how stem cells differentiate as wounds heal.

The field is inherently interdisciplinary. Microscope developers usually hail from biophysics, which is rich in optics research, but cell biologists and chemists also have key roles in the field, with their expertise in handling living material and labelling it non-invasively. Single-cell-imaging labs are found in physics, chemistry and cell-biology departments at universities. Others are affiliated with medical schools, at which imaging experts aid in disease research, diagnostics and drug discovery. Singer and Gratton say that their students and postdocs find careers in both academia and industry.

A single-cell researcher’s academic background isn’t as important as their ability to integrate approaches from various fields. They must apply a quantitative approach to all things biological — rather than just observing cells under the microscope, they must measure how the contents of cells change, for example, or quantify how cellular contents, properties and molecular dynamics vary between cells. And it helps to enjoy tinkering with optical devices.

HIGH-TECH TOYS

High-end fluorescence microscopes cost between US\$300,000 and \$1.5 million, and US federal funding agencies are getting plenty of requests for them. Last year, proposals for microscopes accounted for about half of the grant applications submitted to the biology arm of the Major Research Instrumentation Program at the US National Science Foundation. ►

► The US National Institutes of Health (NIH) Shared Instrumentation Grant Program supports the purchase of research equipment that costs \$100,000–600,000; requests for high-resolution, high-contrast confocal microscopes doubled between 2008 and 2010. And the NIH's High End Instrumentation Program, which supports instruments in the range of \$750,000 to \$2 million, has seen applications for confocal microscopes roughly double every two years since 2005. Price tags are increasing as instruments become more versatile: the average cost of microscopes funded by the German Research Foundation was €300,000 (US\$403,000) five years ago, but is €480,000 now.

Advances in the past decade have brought better probes and microscopes, and improved tools for analysing gene expression. To capitalize on this, the NIH has launched a \$90-million, five-year single-cell-analysis programme to support single-cell imaging and studies on the genomes and other aspects of individual cells. It will award its first grants in 2012 — as long as funds from NIH budgets permit. Other microscope funding programmes include the multi-user equipment grants from the Wellcome Trust in London, which fund microscopes costing up to £1 million (US\$1.6 million), and the Netherlands Organisation for Scientific Research's Investment Subsidy NWO Large programme. Finding out how cells differ is important in disease, says Richard Conroy, programme director at the National Institute of Biomedical Imaging and Bioengineering in Bethesda, Maryland. Not all cancer cells are the same, for example, and it is important to learn which pose a risk to health and which are resistant to treatment.

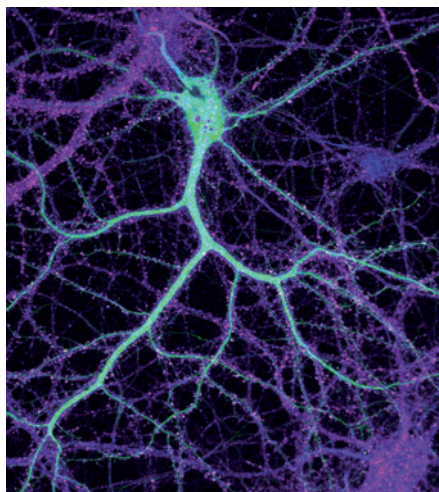


"You want to apply your skills and toolbox to a research problem that will have a future."

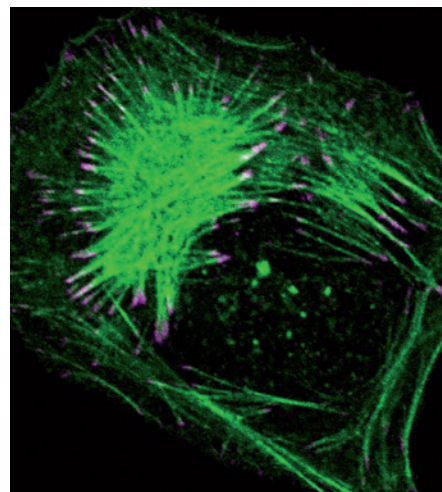
Rick Horwitz

FINE DETAIL

Gratton spends most of his day doling out advice on imaging experiments to the 25 students, technicians and postdocs who work on the 11 custom-made microscopes at his lab. On one occasion, graduate student Chi-Li Chiu told him that an artificial collagen matrix she was handling seemed to be densest around certain breast-cancer cells embedded in it — something that would have been hard to detect before single-cell imaging. To verify the observation, Gratton asked Chiu exactly how dense the matrix became. Chiu used a confocal imaging microscope to take hundreds of snapshots of matrices at different depths. With a computer programme called SimFCS, she analysed the intensity of light reflecting off fibres



Researchers can zoom in on single cells, such as a primary rat hippocampal neuron expressing green fluorescent protein (left), or a connective-tissue cell called a fibroblast (right).



K. LITWA

in the matrix in different places, related those measurements to density and calculated how the living cells deformed the tissue-like matrix.

Chiu's brief task demonstrates the technical and scientific acumen demanded by single-cell imaging — and how it can open up new avenues of research. Gratton now wants to learn how some cancer cells can directly alter the density of tissue. "Is the cell producing more collagen around itself, or is it extending protrusions to grab the collagen and bring it close to the cell?" he asks. "And as these cells are cancerous, does altering the surrounding tissue allow them to migrate and spread?"

Biophysicists often find work in microscope development, which calls for knowledge of the physical properties of lasers. However, they may need contact with a biologist to help them to stay attuned to biological questions and keep cells alive, says Rick Horwitz, a cell biologist at the University of Virginia in Charlottesville, who has a PhD in biophysics. Biophysicists don't necessarily need to move to a cell-biology department, he says, but they should collaborate closely with biologists working in cancer biology, neurobiology, cardiovascular biology or another well-funded field that can be advanced by single-cell imaging. "You want to apply your skills and toolbox to a research problem that will have a future," says Horwitz.

Cell biologists can also become leaders in single-cell imaging — especially now that microscopes and fluorescent probes to tag molecules need not be made from scratch. Manufacturers, including German companies Carl Zeiss in Jena and Leica Microsystems in Wetzlar, and Japanese firms Olympus and Nikon in Tokyo, offer microscope systems built for single-cell and single-molecule imaging, along with the software for analysing the data. Likewise, fluorescent probes can now be purchased from biotechnology companies such as Life Technologies in Carlsbad and BioSearch Technologies in Novato, both in California. With these tools and reagents for sale,

faculty members working in single-cell imaging no longer need the chemical or physical expertise to make their supplies themselves.

Biologists who wish to specialize in single-cell imaging must learn to think quantitatively to design successful experiments, says Paul Selvin, a biophysicist at the University of Illinois at Urbana-Champaign. In his imaging workshops, he has seen biologists who have not grasped this principle. For example, they will predict that a protein will diffuse along a brain-cell membrane and into a synapse, but forget to account for the drag from the fluorescent label that they have attached to the protein, which can slow down or halt diffusion. "It's not very hard to do the calculations to account for drag, but you need to think about the problem in the right terms," says Selvin.

Chemists can contribute to single-cell imaging by developing innovative ways to label molecules. Roger Tsien, a biochemist at the University of California, San Diego, received a share in the 2008 Nobel Prize in Chemistry for his part in discovering the green fluorescent protein, which transformed single-cell analysis by allowing scientists to image the components of living cells non-invasively. And imaging labs often welcome scientists familiar with computer programming, who can develop software to analyse the terabytes of uncompressed image data collected from every experiment.

FINDING A NICHE

Single-cell imaging draws on so many fields that it can be the ideal niche for technology-minded scientists who have struggled to find out where they belong. "There were times when I thought I studied the wrong thing, because I realized I needed to know more biology or more about software development, and I had to learn it on my own," says Christian Hellriegel, an application specialist for Zeiss. "But in the end, I found I had a much wider scope than many people, and that turned out to be a great plus." That scope comes in handy in his job, where

Hellriegel eases communication between university biologists and company physicists who design microscopes. Zeiss and its competitors have been racing to produce high-end microscopes, meaning that there is ever more potential for positions like Hellriegel's.

Hellriegel moved to Zeiss from academia. He has lost the freedom to run his own experiments, but says that not having to write grant applications is well worth any downside. And he is still engaged in his favourite aspect of single-cell imaging — tinkering with microscopes. "One of the things that dawned on me is that there's a lot of thorough science that occurs in industry," he says. "I'm still helping to design equipment for biologists, only now I really have to make sure the equipment works well and is as user-friendly as possible."

Most jobs in single-cell imaging are in university departments and imaging centres. Eric Potma, a physical chemist at the Beckman Laser Institute at the University of California, Irvine, says that researchers should apply to universities at which there are already regular collaborations between imaging centres and medical schools or cell-biology departments. He adds that if the university is not equipped with the necessary high-end instruments and enough space, a large start-up package is essential. "I needed a package of about \$1 million for the kind of work I do, in addition to plenty of space," he explains. His laser system, microscope, reagents, biosafety hood and incubator cost about \$700,000. Another \$300,000 went to support two graduate students and a postdoc for two years. "If you accept lower than what you really need," warns Potma, "your research won't be successful and you may not get tenure." A generous start-up is crucial, Gratton agrees, because it is nearly impossible for an investigator to get a federal grant to purchase high-end imaging equipment early in their career.

Graduate students and postdocs who are interested in single-cell imaging should seek out interdisciplinary labs, advises Singer. "It's great when the people who build the microscopes, make the probes and understand the biology get together in one place," he says. "It's really a great synergy and everything progresses quickly because you can bring a lot more breadth to the problem."

Despite the daunting amount of knowledge required to become an expert in single-cell imaging, those who succeed will be gainfully employed, says Singer. Potma sees the field as part of the progression towards better understanding of how cells contribute to health and disease. "There's still so much left to visualize," he says, "that I don't see an end to this field any time soon." ■

Amy Maxmen is a freelance writer based in New York.

TURNING POINT

Jonathan Cirtain

On 26 September, Jonathan Cirtain, a solar physicist at NASA's Marshall Space Flight Center in Huntsville, Alabama, received a Presidential Early Career Award for Scientists and Engineers.

What made you pursue a scientific career?

When I started college, I went to play American football at the University of Alabama in Tuscaloosa, but an accident in 1993 ended my football career. Lacking motivation, I didn't start college again until 1999 — after my dad made me promise to get a degree and stop wasting my life. He died two weeks later and I kept my promise. I started an undergraduate degree in mathematics and physics, subjects I'd always enjoyed, at the University of Memphis, Tennessee. I wanted to work for NASA or an outfit that supplied NASA with instrumentation. I thought that the best way to achieve that was to get a PhD in physics.

How did you get interested in solar physics?

At a conference of the American Astronomical Society I met Piet Martens, an astrophysicist at Montana State University (MSU) in Bozeman, who told me that his team would be opening a slot to work on NASA's Transition Region and Coronal Explorer (TRACE), which launched in 1998. The telescope took high-resolution images of the Sun's atmosphere, and MSU was involved in mission operations. As a graduate student, I got to operate the scientific payload. I was dropped into the middle of the action.

How did you earn a predoctoral fellowship?

The Sun's atmosphere had always looked fuzzy and out of focus in observations. It turns out that the community was not achieving the best resolution for engineering reasons. During my first year at MSU, I demonstrated a physical model for the Sun's atmosphere using observations from different instruments in space, and showed how imaging instrumentation should be altered to bring the Sun's atmosphere into focus. That caught the attention of Leon Golub at the Harvard-Smithsonian Center for Astrophysics in Cambridge, Massachusetts, and led to a predoctoral fellowship in 2004. Since then, my work has included helping to build the next-generation telescope for TRACE.

How did you accelerate your academic career so quickly?

While calibrating the Hinode X-ray telescope in 2005 at the Marshall Space Flight Center, I got to analyse the data and demonstrate performance. I became one of the few people who



really understood the ins and outs of how the instrument works. I worked on my dissertation and defended it in March 2005. So, in 5 years and 10 months, I went from no undergraduate degree to a PhD.

What are your next big launches?

In June next year, as head of the solar-physics team, I'm going to launch a telescope that will offer a tenfold improvement over any previous extreme ultraviolet instrument. It is called the High Resolution Coronal Imager (Hi-C) and it will have 150-kilometre spatial resolution. A week before I launch that, I will launch the Solar Ultraviolet Magnetograph Instrument (SUMI), which will be used to infer the magnitude and direction of the Sun's magnetic field.

Despite the economic challenges facing NASA, is now a good time to seek a space-related career in the United States?

Perhaps I'm still young enough that I'm naive. The challenges that NASA faces are the ones the nation faces. This is a real economic crisis in the government; I don't think NASA is particularly different from any other agency. However, there is a shortage of US citizens pursuing careers in scientific instrumentation. Most of the graduate students who are interested in this area are foreign nationals. I think that's bad for the nation's scientific competitiveness. Now is as good a time as any to go into space-based astrophysics.

To what do you attribute your career success?

I don't think I'm a genius; I work hard and make the most of opportunities I'm given. I don't know how I would have got this far without my wife, whose support and willingness to move around with me have been extremely important. ■

INTERVIEW BY VIRGINIA GEWIN

GIFTS OF THE MAGI

A career underground.

BY ANATOLY BELILOVSKY

Gregor Samsa awoke one morning from uneasy dreams to find himself transformed into a giant cockroach. The metamorphosis startled him in a casual, momentary way, much as waking up in a strange hotel room might have, as did the darkness into which he woke. He flexed his limbs; with a crackling noise his chrysalis fractured, and morning sunlight flooded his compound eyes. He tried to blink, but although he could not feel his eyelids close, his eyes adjusted to the light much faster than his old ones could; in an instant, the whole room came into sharp, clear focus.

He did not need to dip his head to see each chitinous shard fall spinning to the floor, sparkling in the sunlight. He saw each dust mote in the slanted sunbeams, each flower on the wallpaper, each wrinkle on his bed.

The door inched open. Samsa's daughter poked her head into the room. "Hi Daddy!" she shouted and disappeared. He heard her steps distinctly, and new harmonics in her timbre. He tried to memorize her face as he now saw it, and her voice.

Smells registered next, his wife's familiar scent first of all and, after that, the aroma of dark, strong coffee, laden with sugar, wafted from the kitchen. There was a clinking noise, slower, heavier steps, the door creaked, opened wider —

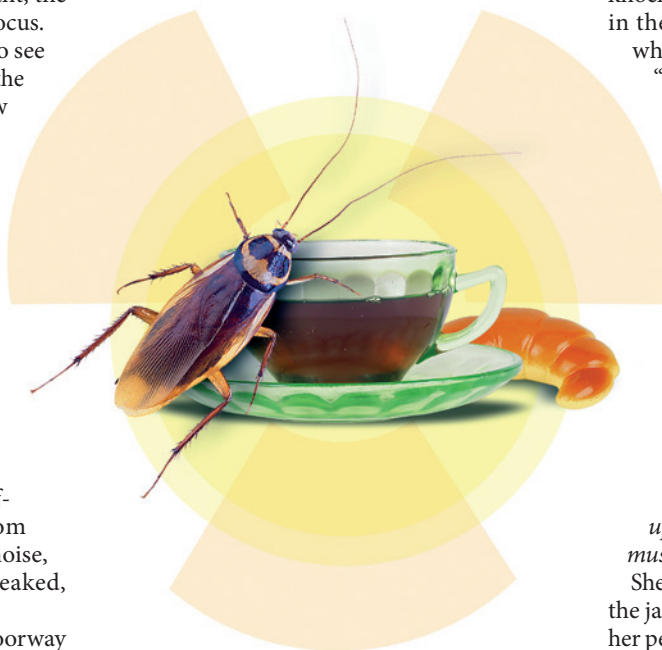
Samsa's wife came through the doorway with the coffee, holding with both hands a half-full bowl. She carried it slowly, placed it in front of him on the floor, sat down on the bed. For a moment, Samsa turned to bring the densest part of his eye towards her. He saw new wrinkles, bloodshot eyes, drying tears. He turned and dipped his proboscis into the coffee, taking her out of focus, but not out of sight.

"I know you can't talk," she said. She looked down, smoothed her hair back, then looked at him again. "I'll talk for both of us. You'll say, 'it's only for a month, you know'." She sniffled. "And I'll think, a whole month. I'll imagine the pipes and the tunnels and the dangers, and you'll tell me that you're now a highly trained professional, and..."

She tried to caress

his carapace; her hands shook, and for a moment her fingers drummed on the unyielding thorax. Her hand recoiled.

"It's the thought of you being underground," she said. "And, honestly, of you not being here. I wish you didn't have to do this. You always used to worry about me flying, and I always told you I'm the best pilot in the sky. I always paid attention to everything, always knew what everyone was doing." Her mouth tightened. "Except you." Her hands wandered as if searching for something familiar: instruments, propulsion controls, handholds; anything.



"Why didn't... Oh hell, I know why you didn't tell me. You wanted to pay off the house by December, so I wouldn't have to take the Saturn run. I didn't tell you I quit, either. I wanted it to be a surprise, that we'd never be apart that long again. So did you. We both love surprises. Remember how I proposed?"

Gregor remembered. A chartered sub-orbital to Paris. Twenty minutes of a ballistic should have been time enough, she'd thought: she would propose, he would accept, then they would make love in zero-G. Except for his nineteen and three-quarter minutes of space sickness. They got it right, finally, in the Royal Suite at Hotel George V, but he never shook off the feeling she'd been left disappointed. *You should see them weightless*, she had said wistfully,

cupping her breasts as she dressed the next morning.

"And you will tell me that this body is damn near indestructible, doesn't even need to breathe or eat," she continued, "and —" her voice cracked — "reactor maintenance pays great." She paused to wipe a tear. "Better than any desk job I could get. Not easy to find work for an ex-spacer with a family to feed," she finished in a whisper.

I should have said something, Samsa thought, *last night. I should have kissed her.*

There was a knock on the door. Samsa's wife smiled. "You raised her right. Always knock first when your parents are together in the bedroom." She stood up. "I know what you'd say if you could," she said.

"You would tell me that you love me, and that you are doing this for us. And I'd say, I love you too. And I'll miss you. Very much. I'll think of nothing else all day."

Yes you will, Samsa thought. *You'll think of flying the way you never thought of it before, just like I never thought of air until I stopped breathing.*

She shambled towards the kitchen with a spacer's graceless gait, but although she'd always been awkward under gravity, for the first time Samsa thought of a long-forgotten line: "*... as if treading upon knife blades so sharp that blood must flow.*"

She stopped in the doorway, one hand on the jamb, turned to face him. He could see her perfectly well without moving his head, but it seemed right to raise it. She swam into full focus again: porous skin, sagging cheeks; but he could see her, too, all of her. She was so beautiful.

"I'm glad we had this talk," she said, willing a smile. "Have a nice day at work, dear."

Samsa nodded. It was the least he could do. And the most. ■

Anatoly Belilovsky was born in what is now Ukraine and learned English from Star Trek reruns. He worked his way through a US college by teaching Russian while majoring in chemistry, and has, for the past 25 years, been a paediatrician in New York, in a practice where English is the fourth most commonly spoken language. He has so far sold SF stories to Andromeda Spaceways, Ideomancer, the Immersion Book of Steampunk and other markets.

JACEY

NATURE.COM
Follow Futures on
Facebook at:
[go.nature.com/mtoodm](https://www.facebook.com/mtoodm)

Fundamental limits of ‘ankylography’ due to dimensional deficiency

ARISING FROM K. S. Raines *et al.* *Nature* **463**, 214–217 (2010)

Raines *et al.*¹ propose a method, which they call ‘ankylography’, for three-dimensional structure determination using single-shot diffractive imaging (SSDI). But the conclusion without limitation that the three-dimensional structure of an object is “in principle encoded into a 2D diffraction pattern on the Ewald sphere” and may be inverted by SSDI is inadequately substantiated and conceptually misleading. Here I point out that SSDI in general suffers from a dimensional deficiency that limits the applicability of ankylography to objects that are small-sized in at least one dimension or that are approximately two-dimensional in some other way.

The rate of reliable information transfer via a spatial or temporal channel is fundamentally limited by the channel capacity, which is determined both by the number of degrees of freedom available therein and by the obtainable signal-to-noise ratio². The resolving powers of telescopes or radio antennas and microscopes including SSDI are all limited in much the same manner. A steep (exponential) price in signal power must be paid to obtain data rates or resolutions significantly beyond those supported by the available number of degrees of freedom^{2,3}. SSDI uses a spatial channel characterized by a linear operator T that projects any real-space amplitude with support $[0, l]^3$ onto the Ewald sphere. It is known^{2,4} that there exist a pair of orthonormal bases $\{u_i\}_{i \geq 1}$ and $\{v_i\}_{i \geq 1}$, called normal modes, and also their associated modal gains $\{\lambda_i(T)\}_{i \geq 1}$, all non-negative and arranged in a non-decreasing order such that $Tu_i = \sqrt{\lambda_i} v_i$, $\forall i \geq 1$. For any predetermined modal cutoff threshold $\varepsilon \in (0, 1)$, $N(T, \varepsilon) = \max\{i : \lambda_i(T) \geq \varepsilon\}$ is the number of usable normal modes, that is, the number of usable degrees of freedom. It has been rigorously proved⁵ that $N(T, \varepsilon) \leq 8l^2 + O(|\log \varepsilon| |\log l| + \log^2 \varepsilon \log^2 l)$, which grows more slowly than the number of unknowns $O(l^3)$ in a general three-dimensional object as its size increases. The insufficiency of the number of useable degrees of freedom would persist even if ε^{-1} grew exponentially as l increased, as long as the exponent grew no faster than $l^{3/2}$: more specifically, so long as $|\log \varepsilon| = O(l^{1.5 - \delta})$, for any fixed small $\delta > 0$.

I emphasize the fundamental nature of the limitation: that single-shot diffraction does not convey sufficient information about the three-dimensional structure of an object, even if the amplitude (instead of intensity) of the diffracted field is sampled continuously and measured directly with no phase ambiguity. Given a practically obtainable signal-to-noise ratio and a measurement accuracy that together determine a threshold ε of modal cutoff, any signal in a linear space spanned by normal modes of orders higher than $N(T, \varepsilon)$ is essentially lost in transmission or attenuated beyond detection. Oversampling and inversion algorithms are irrelevant in this context.

I note that there has been substantial criticism^{6–8} of ankylography even before its publication¹. However, no agreement seems to have been reached on the fundamental problem of dimensional deficiency in ankylography, and an upper bound for the number of degrees of freedom suggested by one critic⁶ has been disputed and its applicability questioned by Raines *et al.*^{1,7,8}, whose response demands a mathematically rigorous proof like the one given here. The present analysis also demonstrates the need to quantify noise and a means of doing so, in order to determine a cutoff threshold based on signal-to-noise ratio, which in turn defines the number of useable degrees of freedom. Although Raines *et al.*^{1,7,8} discuss degrees of freedom, matrix ranks and data sufficiency, they fail to quantify the noise. To say simply that

“Poisson noise was added” to the diffraction patterns does not specify how noise addition was implemented in their numerical simulations, which are the only justification of ‘ankylography’ in the absence of a theoretical base. This vagueness about noise makes it difficult to interpret the reported results and to reconcile the apparent contradiction between the authors’ (ref. 1) and the critic’s (ref. 6) simulation results. Raines *et al.*^{1,8} cite “full rank of matrix” and other numerical results to support the claim that a curved surface of detection and oversampling yielded sufficient information for three-dimensional structure inversion, but they have not responded to the critic’s numerical examples and figures showing rapid decays of singular values⁶.

In summary, SSDI of truly three-dimensional structures does not scale. The applicability of ankylography is limited to objects that are small-sized with respect to the wavelength in at least one dimension or have structures that are essentially two-dimensional in complexity. Such may be the case in the computer tests and preliminary experiment of ref. 1. Raines *et al.*¹ also emphasize certain “physical constraints”, many of which are actually steps of numerical procedures instead of mathematical constraints of model formulation. Incorporating more genuine physical constraints could possibly alleviate the problem of dimensional deficiency. However, that would diminish the generality and appeal of ankylography, and the same feat is arguably achievable by conventional diffractive imaging with a flat image detector.

METHODS

With $F_y F_x$ being a two-dimensional Fourier transform and T_z being a linear operator of z integration⁹, the upper bound on $N(T, \varepsilon)$ follows from the well-known result⁹ that $N(F_x, \varepsilon) = N(F_y, \varepsilon) = 2l + O(|\log \varepsilon| \log l)$, in conjunction with inequalities $N(T, \varepsilon) \leq 2N(T_z F_y F_x, \varepsilon/2) \leq 2N(F_y F_x, \varepsilon/2) \leq 2N(F_y, \varepsilon/2)N(F_x, \varepsilon/2)$, with the second inequality following from an operator inequality¹⁰ $F_x^* F_y^* T_z^* T_z F_y F_x \leq \|T_z\|^2 F_x^* F_y^* F_y F_x$.

Haiqing Wei¹

¹Ambow Research Institute, Ambow Education Group, Beijing 100088, China.

email: davidhwei@yahoo.com

Received 28 January 2010; accepted 21 September 2011.

1. Raines, K. S. *et al.* Three-dimensional structure determination from a single view. *Nature* **463**, 214–217 (2010).
2. Gallager, R. G. *Information Theory and Reliable Communication* (Wiley, 1968).
3. Hansen, R. C. Fundamental limitations in antennas. *Proc. IEEE* **69**, 170–182 (1981).
4. Miller, D. A. B. Spatial channels for communicating with waves between volumes. *Opt. Lett.* **23**, 1645–1647 (1998).
5. Wei, H. Fundamental limits of “ankylography” due to dimensional deficiency. Preprint at <http://arXiv.org/abs/1001.4594v1> (2010).
6. Thibault, P. Feasibility of 3D reconstructions from a single 2D diffraction measurement. Preprint at <http://arXiv.org/abs/0909.1643v2> (2010).
7. Miao, J. Response to “Feasibility of 3D reconstruction from a single 2D diffraction measurement”. Preprint at <http://arXiv.org/abs/0909.3500v1> (2009).
8. Miao, J. & Chen, C.-C. 2nd Response to “Feasibility of 3D reconstruction from a single 2D diffraction measurement”. Preprint at <http://arXiv.org/abs/0910.0272v1> (2009).
9. Landau, H. J. & Widom, H. Eigenvalue distribution of time and frequency limiting. *J. Math. Anal. Appl.* **77**, 469–481 (1980).
10. Bhatia, R. *Perturbation Bounds for Matrix Eigenvalues* (SIAM, 2007).

Competing financial interests: declared none.

doi:10.1038/nature10634

Non-uniqueness and instability of ‘ankylography’

ARISING FROM K. S. Raines *et al. Nature* **463**, 214–217 (2010)

Raines *et al.*¹ have proposed ‘ankylography’ as a three-dimensional tomographic modality with a single monochromatic beam¹. They believe that the three-dimensional structure of an object is “in principle determined by the 2D spherical pattern” that “is sampled at a sufficiently fine scale on the Ewald sphere.” Although they restricted their approach to unspecified “certain circumstances”, they claim potential “broad applications in the physical and life sciences”. Nevertheless, the imaging principles and proposed methodology of ankylography are flawed, bringing into question its applications, which was also pointed out by others, such as in ref. 2.

In the continuous domain, if a significant portion of the Fourier spectrum is unavailable, the recovery of the corresponding object is typically unstable^{3,4}. This type of instability is well known in imaging areas such as X-ray computed tomography and magnetic resonance imaging. For example, X-ray limited-angle tomography theoretically has a unique solution but in practice it yields serious artefacts^{5,6}. In the limiting case, ankylography represents an extreme of limited-angle tomography. With knowledge only of an Ewald sphere, two-dimensional analytic extension is impossible^{5–7}. Given a sufficiently complex structure in the object support, its Fourier spectrum outside the surface of the Ewald sphere can uniquely and stably determine that object, and the solution from data on the surface of the Ewald sphere or another zero-measure surface is generally non-unique and highly unstable.

Let $\Omega \in \mathbb{R}^3$ be an open bounded set. Suppose that $u(\mathbf{r}) \in C_0^4(\Omega)$ is a compactly supported function whose Fourier transform is $\hat{u}(\xi)$, with $\mathbf{r} = (x, y, z)$ and $\xi = (\xi_1, \xi_2, \xi_3)$. Denoting

$$g(\mathbf{r}) = \left(\frac{\partial^4}{\partial x^4} + \frac{\partial^4}{\partial y^4} + \frac{\partial^4}{\partial z^4} + 2 \frac{\partial^4}{\partial x^2 y^2} + 2 \frac{\partial^4}{\partial x^2 z^2} + 2 \frac{\partial^4}{\partial y^2 z^2} + 4R^2 \frac{\partial^2}{\partial x^2} \right) u(\mathbf{r}) \quad (1)$$

we have

$$\begin{aligned} \hat{g}(\xi) &= [(i\xi_1)^4 + (i\xi_2)^4 + (i\xi_3)^4 + 2(i\xi_1)^2(i\xi_2)^2 + 2(i\xi_1)^2(i\xi_3)^2 \\ &\quad + 2(i\xi_2)^2(i\xi_3)^2 + 4R^2(i\xi_1)^2] \hat{u}(\xi) \\ &= (\xi_1^4 + \xi_2^4 + \xi_3^4 + 2\xi_1^2\xi_2^2 + 2\xi_1^2\xi_3^2 + 2\xi_2^2\xi_3^2 - 4R^2\xi_1^2) \hat{u}(\xi) \\ &= [(\xi_1^2 + \xi_2^2 + \xi_3^2)^2 - 4R^2\xi_1^2] \hat{u}(\xi) \\ &= (\xi_1^2 + \xi_2^2 + \xi_3^2 - 2R\xi_1)(\xi_1^2 + \xi_2^2 + \xi_3^2 + 2R\xi_1) \hat{u}(\xi) \\ &= [(\xi_1 - R)^2 + \xi_2^2 + \xi_3^2 - R^2][(\xi_1 + R)^2 + \xi_2^2 + \xi_3^2 - R^2] \hat{u}(\xi) \end{aligned} \quad (2)$$

Equation (2) shows that the Fourier transform of $g(\mathbf{r})$ is zero on the Ewald spheres centred at $(\pm R, 0, 0)$ with radius R . From equation (2), we can construct many counterexamples to prove the non-uniqueness of ankylography.

In the discrete domain, instability of ankylography inherited from the continuous domain will be overwhelming when the number of unknowns is substantial. Ankylography incorporates non-negativity, uniformity and continuity in the hybrid-input-output framework⁸. It allows reconstruction of objects with a small number of degrees of freedom, but fails in our experiments even using an enhanced hybrid-input-output algorithm with total variation minimization^{9,10} when an object matrix is not too small. Specifically, we used a $64 \times 64 \times 64$ Shepp–Logan phantom. The oversampling condition¹ was satisfied. When all the Fourier modulus information was available, an accurate image was reconstructed. However, the reconstruction quality degraded rapidly as the known portion of the Fourier modulus was reduced, and became useless from a single view even with an excellent initial image (Fig. 1a). J. Miao of ref. 1 has confirmed that the current ankylography algorithm does not perform well if an object is symmetric or if the image size is much larger than $15 \times 15 \times 15$. We believe that these limitations are fundamental, and should have been

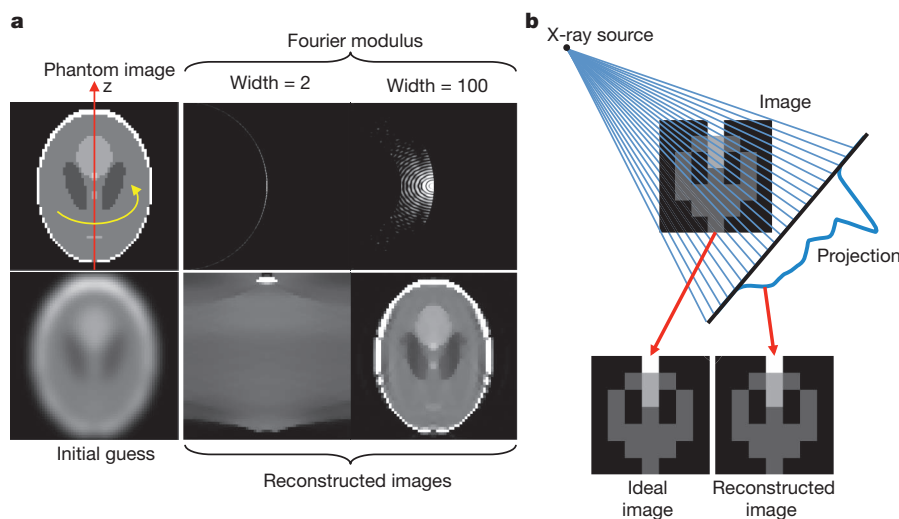


Figure 1 | Numerical studies of ‘ankylography’ and ‘one-view divergent tomography’. **a**, Failure of ankylography with a $64 \times 64 \times 64$ Shepp–Logan phantom (symmetric with respect to the z axis) from noise-free data after 4,000 iterations. We note that the single-view measurement corresponds to the case of width = 1 but here the case of width = 2 is shown for better visibility of data.

The case of width = 100 can be physically studied using an energy-sensitive detector array. **b**, Success of one-view divergent tomography with a 7×7 Shepp–Logan phantom in fan-beam geometry. In this specific setting, the system matrix has a rank (the number of independent measurements) larger than the number of pixels and a reasonable condition number.

mentioned in ref. 1. When the measured Ewald spheres form a layer of substantial thickness (such as the case of width = 100 in Fig. 1a), the reconstruction becomes satisfactory. Data for this type of three-dimensional reconstruction can be simultaneously acquired using an energy-sensitive detector array such as one from the Medipix detector family being developed by CERN¹¹. This approach has an ultrafast imaging speed, and can be referred to as spectrography instead of extended ankylography, because the latter is characterized by its two-dimensional spherical pattern.

The methodological flaw in ankylography may also be illustrated with the example of ‘one-view divergent tomography’ for reconstruction of a two- or three-dimensional structure from a single fan-beam X-ray projection¹²:

$$p(\mathbf{s}, \boldsymbol{\theta}) = \int_0^\infty f(\mathbf{s} + t\boldsymbol{\theta}) dt \quad (3)$$

where \mathbf{s} is the X-ray source position, and $\boldsymbol{\theta}$ is the unit directional vector. Following the same logic as in ref. 1, we could claim that one-view divergent tomography can reconstruct a two- or three-dimensional object using the oversampling strategy. Indeed, it works well for small-size matrices (Fig. 1b). However, it is well known that the solution to this problem is non-unique and unstable, with little utility in the computed tomography field so far.

Ge Wang¹, Hengyong Yu¹, Wenxiang Cong¹ & Alexander Katsevich²

¹SBES Biomedical Imaging Division, VT-WFU School of Biomedical Engineering and Sciences, Virginia Tech, Blacksburg, Virginia 24060, USA.

email: wangg@vt.edu

²Department of Mathematics, University of Central Florida, Orlando, Florida 32816, USA.

Received 26 February 2010; accepted 21 September 2011.

1. Raines, K. S. *et al.* Three-dimensional structure determination from a single view. *Nature* **463**, 214–217 (2010).
2. Thibault, P. Feasibility of 3D reconstruction from a single 2D diffraction measurement. Preprint at <http://arXiv.org/abs/0909.1643v1> (2009).
3. Crimmins, T. R. & Fienup, J. R. Uniqueness of phase retrieval for functions with sufficiently disconnected support. *J. Opt. Soc. Am.* **73**, 218–221 (1983).
4. Foley, J. T. & Butts, R. R. Uniqueness of phase retrieval from intensity measurements. *J. Opt. Soc. Am.* **71**, 1008–1014 (1981).
5. Hamaker, C. *et al.* The divergent beam x-ray transform. *Rocky Mountain J. Math.* **10**, 253–284 (1980).
6. Ye, Y., Yu, H. & Wang, G. Interior tomography: mathematical analysis. In *Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning, and Inverse Problems* (eds Censor, Y., Jiang, M. & Wang, G.) 543–561 (Medical Physics Publishing, 2010).
7. Inouye, T. Image reconstruction with limited angle projection data. *IEEE Trans. Nucl. Sci.* **26**, 2665–2669 (1979).
8. Fienup, J. R. Reconstruction of an object from modulus of its Fourier transform. *Opt. Lett.* **3**, 27–29 (1978).
9. Donoho, D. L. Compressed sensing. *IEEE Trans. Inf. Theory* **52**, 1289–1306 (2006).
10. Yu, H. & Wang, G. Compressed sensing based interior tomography. *Phys. Med. Biol.* **54**, 2791–2805 (2009).
11. Anderson, N. G. *et al.* Spectroscopic (multi-energy) CT distinguishes iodine and barium contrast material in MICE. *Eur. Radiol.* **20**, 2126–2134 (2010).
12. Natterer, F. *The Mathematics of Computerized Tomography* Chs II, IV and VI (John Wiley & Sons, 1986).

Author Contributions G.W. conceptualized and wrote the manuscript. G.W., H.Y. and W.C. identified non-uniqueness of ankylography in simple cases, and A.K. constructed the general argument. H.Y., W.C. and G.W. analysed stability of ankylography and simulated spectrography.

Competing financial interests: declared none.

doi:10.1038/nature10635

FORUM OPTICS

Perfect lenses in focus

Materials that refract light backwards are thought to be required for making super-resolution lenses. An alternative proposal — that conventional, positively refracting media can do the job — has met with controversy. Two experts from either side of the debate lay out their views on the matter.

Positive thinking

TOMÁŠ TYC

In 2000, Pendry showed¹ that a slab of material that bends light at a negative angle can work as a lens with the ability to resolve details much smaller than the wavelength of light. This is due to the fact that, unlike conventional lenses, which refract light at a positive angle (Fig. 1), this device transfers not only the propagating (long-range) waves of light from an object to its image, but also the object's evanescent waves — short-range light that carries smallest-scale information about the object. However, such a perfect lens, although based on a neat idea, has some serious drawbacks. For example, it turns out² that, for fundamental reasons, negative refraction is always connected with light absorption, and such absorption destroys the super-resolution ability of the lens. Moreover, perfect lenses based on negative refraction are difficult to manufacture and can work only in narrow bands of the electromagnetic spectrum.

A natural question to ask, therefore, is whether super-resolution lenses could be achieved by using materials that have a purely positive refractive index. In my opinion, the answer is definitely yes. As Leonhardt has shown³ by analytical calculations, Maxwell's fish eye⁴ — a prototype of a positively refracting perfect lens — can provide imaging that has, in principle, unlimited resolution. This theoretical prediction was confirmed experimentally by Ma, Leonhardt and colleagues⁵, who showed that images of two sources of microwave radiation (used instead of light), separated by one-fifth of the radiation's wavelength, could be clearly distinguished.

However, to achieve such super-resolution by Maxwell's fish eye, an outlet (drain) is required so that the radiation reaching the point of image formation can be absorbed or otherwise extracted. And it is this feature that is at the root of the controversy surrounding the issue of using positive refraction to make perfect lenses. Using a drain is not a problem, however, because the very reason for imaging

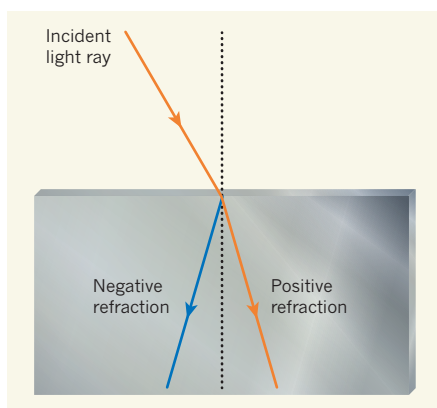


Figure 1 | Positive versus negative refraction. Unlike conventional materials, which refract incident light at a positive angle, artificially engineered materials that have a negative refractive index bend light at a negative angle.

is to record the image on some photosensitive medium, which naturally provides the outlet.

There are some similarities between imaging by Maxwell's fish eye and by time-reversal mirrors^{6,7}. In both cases, the object's electromagnetic waves converge at the image point from all directions to create a subwavelength-resolution image. To produce a time-reversal mirror, the electromagnetic field must be recorded, inverted in time and then re-emitted using a complicated set-up involving active elements (additional sources of radiation). By contrast, Maxwell's fish eye and other positively refracting perfect lenses form the converging waves naturally, without the need for active elements or field recording.

Although the theoretical³ and experimental results⁵ are promising, there are still many unresolved challenges relating to super-resolution with positive refraction. Probably the most exciting one is how to apply Maxwell's fish eye and other perfect lenses in microscopy or nanolithography — the two fields that these devices are most likely to revolutionize.

Tomáš Tyc is in the Faculties of Science and of Informatics, Masaryk University, 61137 Brno, Czech Republic.
e-mail: tomtyc@physics.muni.cz

No drain, no gain

XIANG ZHANG

I take issue with Leonhardt and colleagues' claim^{3,5} that Maxwell's fish eye is a perfect lens. Maxwell's fish eye, proposed⁴ more than 150 years ago, is subject to a diffraction limit: it cannot resolve any feature smaller than a fraction of the wavelength of the light being used.

Over the past decade, negative-index metamaterials, which are made of artificially structured composites, have been used as a means to overcome the diffraction limit and to make a perfect lens by focusing all wave components of light emitted or scattered from the object^{1,8}. The key to such a perfect lens is its very ability to restore the smallest features of the object by enhancing the evanescent waves, which often decay in space.

Leonhardt and colleagues argued^{3,5} that a metal-coated Maxwell's fish eye, which is made of a positive-index material, can also act as a perfect lens by collecting all wave components. Their trick for attaining a perfect lens was to place an additional optical active element (the drain) exactly where the object's image is formed.

The problem with this approach lies in the physical interpretation of the imaging resolution beyond the diffraction limit. An image formed using the drain-assisted fish-eye system involves electromagnetic waves not only from the object but also from a new source — the drain. The image is therefore no longer an intrinsic property of the fish eye itself. It was shown^{9,10} that removal of the drain destroys the sub-diffractional object details, resulting in a diffraction-limited image. It is therefore not justified to claim that a general positive-refracting material can make a perfect lens.

Placing the drain at the image position supplies, through an electromagnetic field induced in the fish eye, the time-reversed form of the object's electromagnetic waves, and the superposition of the time-reversed waves yields an apparently perfect image. The device thus falls within well-known super-resolution

image schemes based on time reversal⁷.

The drain-assisted perfect lens is, however, an interesting use of Maxwell's fish eye, and it may offer opportunities from operations known as non-Euclidean optical transformations. Conventionally, the lens is an independent device that is separated from the object and its image. By contrast, with the fish-eye lens, both the object and image are embedded in it. How the embedded object and image affect the lens and its functions remains to be investigated. For example, displacement of the space inside the lens by an object of finite

size can significantly alter how the refractive index varies across the lens and therefore the lens's optical functions. What's more, detecting the image from inside the fish eye can be a challenge for practical applications. Nevertheless, the 'entangled' or integrated approach of an object–lens–image with a drain is an idea worth exploring. ■

Xiang Zhang is in the NSF Nanoscale Science and Engineering Center, University of California, Berkeley, Berkeley, California 94720–1740, USA.

e-mail: xiang@berkeley.edu

1. Pendry, J. B. *Phys. Rev. Lett.* **85**, 3966–3969 (2000).
2. Stockman, M. I. *Phys. Rev. Lett.* **98**, 177404 (2007).
3. Leonhardt, U. *New J. Phys.* **11**, 093040 (2009).
4. Maxwell, J. C. *Camb. Dublin Math. J.* **8**, 188 (1854).
5. Ma, Y. G., Sahebdivan, S., Ong, C. K., Tyc, T. & Leonhardt, U. *New J. Phys.* **13**, 033016 (2011).
6. Lerosey, G., de Rosny, J., Tourin, A. & Fink, M. *Science* **315**, 1120–1122 (2007).
7. de Rosny, J. & Fink, M. *Phys. Rev. Lett.* **89**, 124301 (2002).
8. Fang, N., Lee, H., Sun, C. & Zhang, X. *Science* **308**, 534–537 (2005).
9. Blaikie, R. J. *New J. Phys.* **12**, 058001 (2010).
10. Merlin, R. *Phys. Rev. A* **82**, 057801 (2010).

EVOLUTION

Big brains explained

The expensive-tissue hypothesis proposes that brain enlargement during human evolution was offset by a reduced gut size. The finding that the typical trade-off in mammals is between brain size and fat reserves suggests otherwise. SEE LETTER P.91

RICHARD POTTS

Brain enlargement is one of the more conspicuous aspects of human evolutionary history. Although the benefits of a large brain seem obvious, the correlated survival and metabolic costs are immense and define much about the human condition. In *Homo sapiens*, giving birth to big-brained babies is risky. Furthermore, the energy consumed by the brain forms roughly 65% of a baby's total consumption and no less than 20–25% of an adult's, even though brain tissue accounts for only 2% of adult body mass.

An enduring question, then, is how the energetic costs of evolving a larger brain were overcome, eventually enabling a threefold increase in volume in the transition from the hominin *Australopithecus* to *H. sapiens* (Fig. 1). On page 91 of this issue, Navarrete *et al.*¹ address this problem by presenting an impressive data set comparing mammalian brain size with the mass of visceral organs and fat deposits. Their results challenge the compelling idea that brain enlargement in *Homo* and other mammals can be 'financed' by a reduction in gut size.

The idea contested by Navarrete *et al.* is known as the expensive-tissue hypothesis². It argues that the gastrointestinal tract is the only energetically costly organ system in humans and other primates that correlates negatively with brain size. Furthermore, a reduced gut is characteristic of primates that have high-quality diets. Because access to substantial quantities of meat and other new food resources has improved the quality of the diet of ancestral humans over the past 2.5 million years, this is expected to have allowed the human gut to become smaller. A key part of the expensive-tissue hypothesis, therefore, is that the costs of

brain expansion in *Homo* were covered by this reduction in gut size.

According to the opposing view now offered by Navarrete *et al.*, the gut–brain trade-off should be replaced by a fat–brain trade-off. The authors took on the ambitious task of dissecting and measuring fat tissue mass in 100 species of mammals. They discovered that brain size correlates negatively with the amount of body fat in most mammals, but not with the mass of the gut, liver or any other tissue that has been proposed to be energetically expensive in mammals.

The lack of a negative relationship between expensive tissues and brain size across many

mammalian groups led the authors to dispute the idea that human brain enlargement was paid for by energy savings associated with a reduced gastrointestinal tract. Instead, they suggest that increasing body fat and brain size are complementary strategies for warding off starvation. In other words, an organism's capacity to store body fat, which is a relatively inexpensive way to buffer food scarcity, can be reduced in lineages in which a bigger brain allows better-quality food intake or lowers the energetic costs of other life functions.

But what about primates? And how does this finding¹ help to explain the particular case of *H. sapiens*? Navarrete *et al.* did not observe a reciprocal relationship between brain size and fat reserves in primate species. They attribute this result to the fact that the primates they studied were captive animals — and thus not truly representative of primates in their natural environment — and to the diverse ways in which primates store fat, among other factors. As for *H. sapiens*, our species is unusual in having not only a large brain, but also hefty fat deposits — a dual strategy for combating starvation that was apparently beneficial

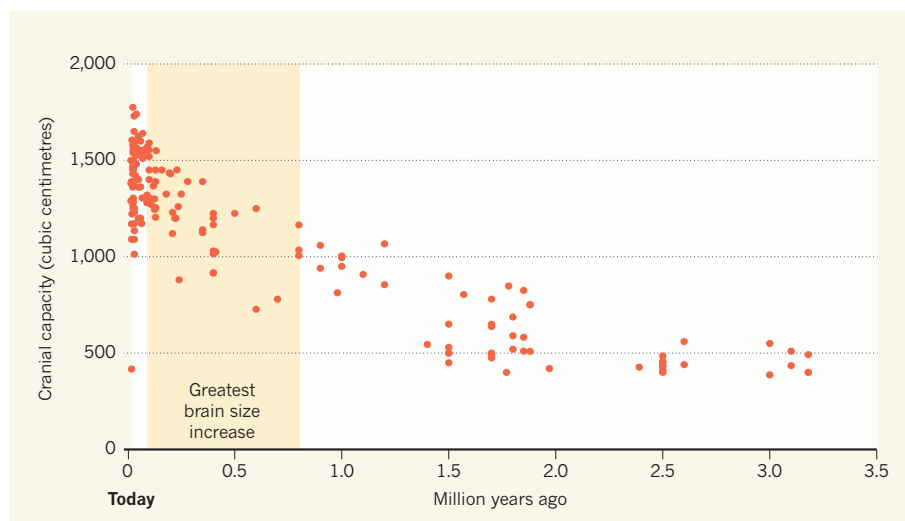


Figure 1 | Human brain expansion. The cranial capacities of hominin fossils illustrate an increase in brain size, largely within the genus *Homo*, over the past 2 million years. If *Homo floresiensis* (bottom-left data point) is omitted as an outlier, the data show that more than 50% of the increase occurred between 800,000 and 200,000 years ago. This suggests that the processes and pathways that caused brain expansion in *Homo* were concentrated in this period. (Cranial-capacity data from refs 3–5.)

in the variable and unpredictable habitats in which *Homo* evolved. Whereas most mammalian species employ one or the other strategy, the evolution of both in *Homo*, according to Navarrete *et al.*, depended on the evolution of energetically efficient bipedal locomotion. Lowering the energy costs of locomotion early on in evolution might have allowed humans to evolve fat-storage capacity along with an expensive brain.

The authors' study¹ admirably draws attention to several other factors relevant to brain expansion in *Homo*. The use of tools, cooking and advances in foraging technologies, for example, would have improved the food supply for *Homo* and so stabilized energy intake. Brain expansion was also helped by the development of behaviour such as giving resources to other individuals — a peculiar dimension of human social interactions that comes with expectations that adults will share and eat food together and cooperate in caring for the young. Such behaviour might have reduced the amount of energy expended by reproductive females and dependent offspring, thereby allowing an increase in brain size and birth rate, along with a slower pace of development for infants.

When considering the above suggestions, one is confronted by the unique outcomes of human evolution. It is good science to seek biological explanations for specific evolutionary features by using principles that are applicable across the widest range of species. Biological principles have, however, led to uncommon outcomes in humans. Compared with other mammals, human carers make an enormous investment in offspring, and they require atypically cooperative efforts to meet the time and energy demands of an infant's prolonged maturation. A keen acquisitiveness of energy-rich resources is another defining human quality, as this is imperative for feeding our hungry brain. These and other remarkable outcomes of general evolutionary processes have led to proposed explanations for the explosion in human brain size — including the expensive-tissue hypothesis — that fall 'outside the box' of general mammalian rules. Such explanations are arguably not ideal, but are acceptable as long as they are based on solid empirical data and comparative biological reasoning.

Navarrete and colleagues, however, pursue an explanation that applies generally across mammals — an approach that is both refreshing and provocative. Inevitably, some will ask: could it be that gut-size reduction helped pay for brain enlargement in our genus, even if it did not do so in most other mammals? It is impossible to rule this out, but the authors' hypothesis¹ does neatly tie together many facets of human evolutionary history, without invoking a gut-brain trade-off. Fat now looms large in explanations of human evolution, and will probably feed debates for years to come. ■

Richard Potts is in the *Human Origins Program, National Museum of Natural History, Smithsonian Institution, Washington DC 20560, USA.*
e-mail: pottsr@si.edu

1. Navarrete, A., van Schaik, C. P. & Isler, K.

- Nature* **480**, 91–93 (2011).
2. Aiello, L. C. & Wheeler, P. *Curr. Anthropol.* **36**, 199–221 (1995).
3. Holloway, R. L., Broadfield, D. C. & Yuan, M. S. *The Human Fossil Record* Vol. 3 (Wiley, 2004).
4. Falk, D. *et al. Proc. Natl Acad. Sci. USA* **104**, 2513–2518 (2007).
5. Carlson, K. J. *et al. Science* **333**, 1402–1407 (2011).

REGENERATIVE MEDICINE

Organ recital in a dish

It is difficult to recapitulate organ development *in vitro*, especially when interactions between tissues are essential. Nonetheless, researchers have now achieved this for the pituitary gland. SEE ARTICLE P.57

KARINE RIZZOTI & ROBIN LOVELL-BADGE

Suga *et al.*¹ report *in vitro* generation of a functional embryonic organ from mouse embryonic stem cells on page 57 of this issue. They have induced formation of Rathke's pouch — a precursor structure to the pituitary gland — from which hormone-secreting cells can be made to differentiate. This work is a notable advance towards studying pituitary development and for better management of pituitary-hormone deficits through regenerative medicine. It also illustrates that inductive processes — whereby tissues interact to form organs — can be triggered in a controlled and largely autonomous manner, which bodes well for the prospect of developing even more complex organs in the lab.

In vertebrates, the pituitary — under the control of the hypothalamus in the brain — regulates essential physiological processes, including growth, puberty and reproduction. In humans, poor functioning of this organ can be congenital or acquired later in life, and affects one in 3,000–4,000 live births. Although hormone-replacement therapy can be effective in these patients, it has its disadvantages, not least that it cannot reproduce normal hormone-secretion patterns².

In the early embryo, the tissues that will form the hypothalamus and Rathke's pouch are adjacent, because they form from the most anterior (rostral) neuroepithelium and the ectoderm cell layers, respectively (Box 1). Although both structures are then displaced posteriorly, they remain in intimate contact: the presumptive hypothalamus induces formation of Rathke's pouch in the developing embryo and regulates pituitary-hormone production after birth³.

Differentiation of rostral hypothalamus from embryonic stem-cell aggregates has been reported previously⁴. Some of the same authors reasoned¹ that induction of an even more anterior feature in these aggregates might allow for the concomitant development of

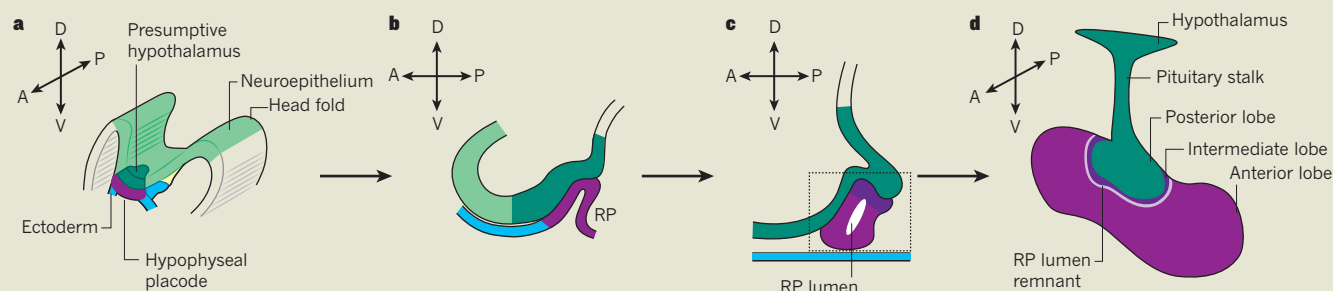
rostral ectoderm along with neuroepithelium — a prerequisite for pituitary development. To see whether this was the case, they obtained floating aggregates in which an external layer of rostral-like ectoderm covered an inner layer of hypothalamus-like neuroepithelium. They then treated the aggregates with activators of the signalling molecule Sonic Hedgehog, a protein that is required for proper development of Rathke's pouch in mice⁵. Suga *et al.* observed spectacular formation of invaginated hollowed vesicles that were strikingly reminiscent of Rathke's pouch *in vivo*, not only in shape and location, but also in the molecular markers they expressed (see Fig. 1 of the paper¹).

It is not surprising that factors required for Rathke's-pouch formation *in vivo* (including Sonic Hedgehog, bone morphogenetic proteins and fibroblast growth factors) are also needed for pouch formation *in vitro*. What was unexpected, however, was such a degree of organization in this simple culture system. *In vivo*, morphogenetic movements (development of the head folds and neural-tube closure) affect the final location of Rathke's pouch. Yet, in Suga and colleagues' experimental system, similar movements, at least for the pouches, were recapitulated outside the 'head' context. It is also noteworthy that, in contrast to the *in vivo* situation, these aggregates did not contain mesodermal or neural-crest tissues, despite data suggesting that at least the mesoderm⁶ is required for Rathke's-pouch development.

The authors¹ also find that inhibition of the Notch signalling pathway — which is involved in cell-fate decisions and, in many systems, in progenitor maintenance versus differentiation — is sufficient to induce differentiation of corticotrophs, the endocrine cells of the anterior pituitary lobe that secrete adrenocorticotrophic hormone. This observation nicely recapitulates previous *in vivo* data^{7,8} obtained from Notch mutant mice. Corticotrophs can also form spontaneously *ex vivo*,

BOX 1

Development of the pituitary and hypothalamus



The ectoderm is the outermost layer of the embryo, and gives rise to the epidermal layer of skin and parts of its associated structures, such as the ears and eyes.

a, During early embryonic development in vertebrates, a localized thickening called the hypophyseal placode forms at the midline of the rostral ectoderm (blue), adjacent to the part of the central nervous system that is the presumptive hypothalamus. This occurs as the head folds begin to take

shape, but, because they grow rapidly compared with the midline region, the early rostral regions are displaced posteriorly.

b, Later in development, the hypophyseal placode remains in close contact with the presumptive hypothalamus, but changes shape as it is pulled upwards, towards the overlying neuroepithelium, to form Rathke's pouch (RP). **c**, By mid-gestation, this simple epithelial invagination separates from the underlying ectoderm to form the definitive

Rathke's pouch. **d**, Subsequent cell proliferation and differentiation events allow formation of the mature anterior pituitary lobe, which contains most of the pituitary's endocrine cell types. Other pituitary components include the intermediate lobe, separated from the anterior lobe by the remnant of the Rathke's-pouch lumen; the posterior lobe; and the pituitary stalk, which connects the gland to the hypothalamus. A, anterior; P, posterior; D, dorsal; V, ventral.

when Rathke's pouch is dissected out and maintained in culture^{9,10}. As all endocrine cell types begin to differentiate during the same period of development¹¹, *ex vivo* and *in vitro* corticotroph differentiation probably reflects a permissive context for this particular cell type at this early time point, rather than a corticotroph-specific inductive event. Indeed, Suga *et al.* report differentiation, albeit less efficient, of all other anterior-lobe endocrine cell types. This is consistent with *in vitro* differentiation of all endocrine cell types from cellular spheres derived from adult pituitary stem and/or progenitor cells².

The authors convincingly show that their *in vitro*-generated corticotroph-containing aggregates are functional when grafted into mice in which the pituitary has been ablated. A previous study¹² also showed that progenitor cells from adult pituitary could differentiate *in vivo*, in vascularized microchambers. However, low cell yield and low differentiation efficiency were limiting factors for the use of cells obtained in this way in regenerative medicine. By contrast, Suga *et al.* had access to an unlimited supply of cells to obtain endocrine cells *in vitro*. Theoretically, this means that endocrine cells may also be obtainable from patient-derived induced pluripotent stem cells, which would offer a real improvement for the management of pituitary-hormone deficits.

More work is required to understand similarities and differences between *in vivo* and *in vitro* development of Rathke's pouch,

and to improve endocrine-cell differentiation — perhaps by including supporting tissues and vasculature, as the authors¹ suggest. Obtaining adult cell types from embryonic stem cells *in vitro* has proved problematic¹³. It should therefore also be investigated whether the embryonic-like pituitary cells that Suga *et al.* generated are fully mature.

Within the mature anterior pituitary lobe, the various cell types are linked together in complex networks that are thought to help coordinate hormone release in response to cues from the hypothalamus or the periphery¹⁴. As part of this response, cell mobility and the organization of these networks can change¹⁵. Is any of this mimicked in Suga and co-authors' *in vitro*-derived pituitaries? Plasticity of the pituitary — whereby the relative proportions of different hormone-secreting cell types can change in response to events such as puberty, pregnancy and lactation — may also depend, in part, on populations of stem cells and their progenitor cells within the gland². It will be interesting to see whether these cells are present in pituitaries generated in the lab. If so, it may be easier to study these cells' origin and behaviour *in vitro*.

Many mutations can cause pituitary defects. But it is not always clear whether these are intrinsic to specific cell types and aspects of pituitary development or due to secondary constraints, imposed by abnormalities elsewhere that compromise inductive interactions. Suga and co-workers' method should allow the question of how the various pituitary

defects develop to be addressed: it can be combined with the powerful technique of live-tissue imaging, which is difficult to apply to developing tissues *in vivo*.

In recent years, the field of tissue engineering has made considerable progress, especially with the use of natural and artificial matrices that can be seeded with tissue-specific stem cells to create various organs^{16,17}. But perhaps — as this paper¹ shows — early embryonic tissues already know what to do. Indeed, it has long been known that teratoma tumours, derived from pluripotent stem cells, can contain a complex mixture of tissues (such as teeth, skin and gut tissue), albeit in a haphazard manner. What Suga *et al.* have achieved is to harness this ability to obtain a robust and directed system. Could it also work with similar types of inductive event, such as those that lead to more complex organs, including the lungs, liver and pancreas¹⁸? ■

Karine Rizzoti and Robin Lovell-Badge
are at the MRC National Institute for Medical Research, London NW7 1AA, UK.
e-mail: rlovell@nimr.mrc.ac.uk

1. Suga, H. *et al.* *Nature* **480**, 57–62 (2011).
2. Castinetti, F., Davis, S. W., Brue, T. & Camper, S. A. *Endocr. Rev.* **32**, 453–471 (2011).
3. Kelberman, D., Rizzoti, K., Lovell-Badge, R., Robinson, I. C. A. F. & Dattani, M. T. *Endocr. Rev.* **30**, 790–829 (2009).
4. Wataya, T. *et al.* *Proc. Natl Acad. Sci. USA* **105**, 11796–11801 (2008).
5. Treier, M. *et al.* *Development* **128**, 377–386 (2001).
6. Gleiberman, A. S., Fedtsova, N. G. & Rosenfeld, M. G. *Dev. Biol.* **213**, 340–353 (1999).

7. Zhu, X. *et al.* *Genes Dev.* **20**, 2739–2753 (2006).
8. Kita, A. *et al.* *Mol. Endocrinol.* **21**, 1458–1466 (2007).
9. Ericson, J., Norlin, S., Jessell, T. M. & Edlund, T. *Development* **125**, 1005–1015 (1998).
10. Treier, M. *et al.* *Genes Dev.* **12**, 1691–1704 (1998).
11. Davis, S. W., Mortensen, A. H. & Camper, S. A. *Dev. Biol.* **352**, 215–227 (2011).
12. Lepore, D. A. *et al.* *Stem Cells* **25**, 1730–1736 (2007).
13. Murry, C. E. & Keller, G. *Cell* **132**, 661–680 (2008).
14. Lafont, C. *et al.* *Proc. Natl Acad. Sci. USA* **107**, 4465–4470 (2010).
15. Schaeffer, M. *et al.* *Endocrinology* <http://dx.doi.org/10.1210/en.2011-1430> (2011).
16. Griffith, L. G. & Swartz, M. A. *Nature Rev. Mol. Cell Biol.* **7**, 211–224 (2006).
17. Orlando, G. *et al.* *Transpl. Int.* **24**, 223–232 (2011).
18. Green, M. D. *et al.* *Nature Biotechnol.* **29**, 267–272 (2011).

PHYSIOLOGY

Immune cells fuel the fire

Regulation of body temperature by the nervous system is essential for physiological function in both health and disease. The immune system also seems to have a crucial role in this process. SEE LETTER P.104

ANDREW J. WHITTLE
& ANTONIO VIDAL-PUIG

All mammals, including humans, are homeotherms — they maintain a constant body temperature regardless of their environment. This essential task is performed by brown fat tissue^{1–3}. Whereas white fat is an energy storage depot, brown fat is equipped with a specialized form of mitochondria, cellular powerhouses, that can convert stored fat directly into heat through the process of thermogenesis⁴. To avoid energy wastage, this happens only when neurons connected to brown fat release chemical messengers called catecholamines. But neurons do not seem to be the only regulators of thermogenesis. On page 104 of this issue, Nguyen *et al.*⁵ show that a subset of immune-system cells called macrophages is also essential for heat production in brown fat*.

Thermogenesis is fundamental to the ability of mammals to balance their energy requirements with their nutritional stores. In rodents, the effectiveness of this process has implications or how easily the animals become obese, and there is increasing evidence for a similar relationship between thermogenesis and body-weight regulation in humans². The ability of white fat tissue to release lipids into the bloodstream as required, or remove them from it, is equally crucial. Humans in whom this process is disturbed through genetic defects become extremely ill: fat is inappropriately deposited in organs

*This article and the paper⁵ under discussion were published online on 20 November 2011.

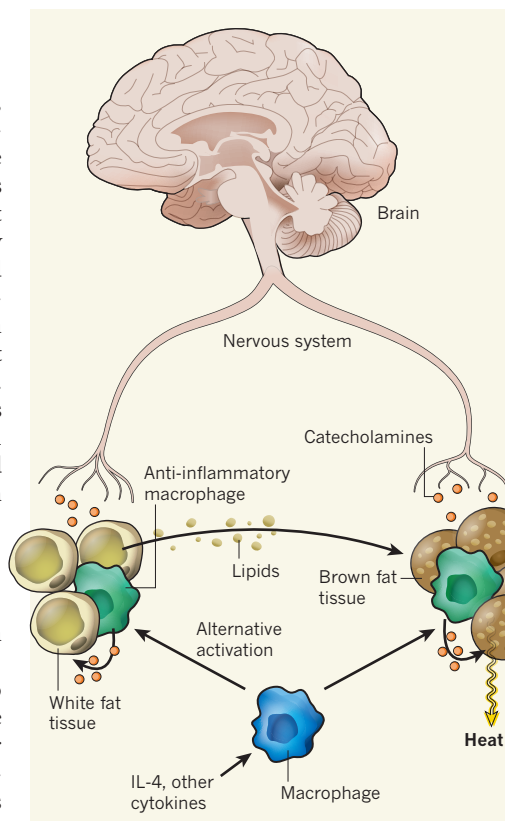


Figure 1 | Role of macrophages in thermogenesis.

In response to a reduced environmental temperature, the brain sends chemical signals (catecholamines) to white and brown fat tissues. Catecholamines activate brown fat to generate heat. The source of energy for heat production is lipids that are released by white fat in response to catecholamines and that reach brown fat through the bloodstream. Nguyen *et al.*⁵ report that IL-4, and perhaps other cytokines including catecholamines themselves, drive alternative (anti-inflammatory) activation of macrophages in both forms of fat tissue. The activated macrophages also secrete catecholamines to enhance and sustain the thermogenic response.

such as the liver, and in muscle, leading to diabetes^{6,7}.

Macrophages are a frontline defence against anything that should not be present in the body. Being highly mobile, they infiltrate almost every tissue to consume and dispose of material that might be damaging. To fight pathogens, macrophages are transformed into pro-inflammatory machines that secrete catecholamines. Along with cytokine proteins, catecholamines regulate the intensity of the immune response at the site of infection⁸. However, macrophages also exist in alternatively activated, anti-inflammatory forms that have a wide range of physiological roles. Nguyen *et al.* explore the regulatory role of alternatively activated macrophages in one such physiological task — thermogenesis.

To induce thermogenesis, the authors⁵ placed mice in a cold environment and found that the macrophages in the animals' brown and white fat underwent a clear shift towards the anti-inflammatory form. Removal of the gene encoding the cytokine protein IL-4 or that encoding the IL-4 receptor from macrophages — thereby preventing them from adopting the anti-inflammatory form — resulted in striking thermogenic defects. The mice could no longer effectively generate heat in their brown fat; they could not maintain their core body temperature in the cold; and they would probably have died after prolonged exposure to low temperatures. These defects were not solely caused by a problem in brown fat: white fat tissue did not respond to nervous stimulation by releasing lipids into the bloodstream and so did not provide fuel for thermogenesis in the brown fat (Fig. 1).

In obese people, macrophages in white fat tend to show increased pro-inflammatory activity⁹ and have been linked to insulin resistance and diabetes. In light of Nguyen and colleagues' data, it might be that pro-inflammatory macrophages cause a detrimental response in obese states in which the fat tissue is under increasing pressure to expand. As fat cells become overburdened and begin to die, pro-inflammatory macrophages would be expected to increase in number to clean up the debris. This could limit the number of macrophages available for transformation into anti-inflammatory forms. In the absence of sufficient anti-inflammatory macrophages, the white fat may not efficiently respond to signals from the central nervous system, creating a state of dysregulated lipid release¹⁰ and metabolic inflexibility.

Beyond the direct implications of these findings⁵ for energy balance lie far-reaching issues for the entire field of animal research. Nguyen and co-authors' observations suggest that, in some settings, the central nervous system relies heavily on macrophages to mediate the appropriate peripheral response to normal physiological demands. But in animal studies, macrophages are often

manipulated to create a range of models from those for Alzheimer's disease to HIV infection. Therefore, any published study that is based on the manipulation of macrophages may need to be re-examined, because scientists should ask whether the effects they observed were the result of direct manipulation of the immune system or a result of secondary alterations in the activity of the central nervous system.

Addressing whether such genetic manipulations alter how the mouse brain 'perceives' its environment is beyond the scope of the present work. Furthermore, Nguyen *et al.* do not explore the proportional contribution of macrophages to nervous-system activity. What they do show, however, is that alternatively activated macrophages are key to how the body handles and burns its fat stores. In that respect, this specialized, widely distributed group of cells could represent a novel target for therapies for obesity and other fat-storage disorders. ■

Andrew J. Whittle and Antonio Vidal-Puig
are in the Department of Clinical Biochemistry, University of Cambridge Metabolic Research Laboratories, Institute of Metabolic Science, NIHR Cambridge Biomedical Research Centre, Addenbrooke's Hospital, Cambridge CB2 0QQ, UK.
e-mails: ajw232@medschl.cam.ac.uk; ajv22@cam.ac.uk

1. Cypess, A. M. *et al.* *N. Engl. J. Med.* **360**, 1509–1517 (2009).
2. Ouellet, V. *et al.* *J. Clin. Endocrinol. Metab.* **96**, 192–199 (2011).
3. Saito, M. *et al.* *Diabetes* **58**, 1526–1531 (2009).
4. Cannon, B. & Nedergaard, J. *Physiol. Rev.* **84**, 277–359 (2004).
5. Nguyen, K. D. *et al.* *Nature* **480**, 104–108 (2011).
6. Gandotra, S. *et al.* *J. Biol. Chem.* **286**, 34998–35006 (2011).
7. Huang-Doran, I. *et al.* *J. Endocrinol.* **207**, 245–255 (2010).
8. Flierl, M. A. *et al.* *Nature* **449**, 721–725 (2007).
9. Weisberg, S. P. *et al.* *J. Clin. Invest.* **112**, 1796–1808 (2003).
10. Prieur, X. *et al.* *Diabetes* **60**, 797–809 (2011).

ASTROPHYSICS

A puzzling γ -ray burst

The population of γ -ray bursts is sometimes enriched by perplexing events that challenge established theoretical models. Two teams get to grips with understanding one such event — the 'Christmas' burst. [SEE LETTERS P.69 & P.72](#)

ENRICO COSTA

Gamma-ray bursts strike Earth's atmosphere from random directions in space about twice a day. In 1997, the discovery¹ of the emission, or afterglow, that follows the prompt release of γ -rays placed their progenitors in remote galaxies. The mechanisms put forward to explain the bursts and their afterglows are complex and differ from one to the next. Any theory for the origin of γ -ray bursts must account for both the commonalities and the differences between them. Writing in this issue, Campana *et al.*² (page 69) and Thöne *et al.*³ (page 72) use two completely different models to explain the especially unusual 'Christmas' γ -ray burst, so dubbed because it occurred on Christmas Day 2010.

Although each one is a unique event, γ -ray bursts (GRBs) can be roughly divided into two classes — long and short — according to the duration of the prompt emission. Long GRBs, which last from 2 seconds to a few minutes, are thought to originate from hypernovae, an unusual class of supernova. They involve the collapse of a massive star (a hundred times more massive than the Sun) that not only explodes as a supernova, but also generates two opposing high-speed jets of particles, leaving behind a black hole. If one of the jets happens

to point towards an observer, a bright emission — the prompt γ -ray release — is detected.

The GRB afterglow, which follows minutes after the initial burst and fades away over the course of weeks or months, is emitted at various wavelengths and is caused by the impact of the jet on the surrounding interstellar medium. Once the GRB fades away, the faint galaxy that hosts the collapsing star can be identified, and the distance of the GRB from Earth measured from the galaxy's spectrum. Short GRBs, which typically last less than 2 seconds, are better explained by two neutron stars bound in a tight binary system. The stars produce gravitational waves and eventually merge to form a black hole, producing the GRB's jets and afterglow in the process of merging.

As GRB research has progressed, established models have been challenged by observations, even of a single GRB. It is conceivable that the GRB population has been 'polluted' — or, alternatively, enriched — by a few puzzling events that mimic GRBs but have a completely different origin. One such event is the Christmas GRB, which was detected by the Swift satellite and is known technically as GRB 101225A. The Christmas GRB's γ -ray emission was exceptionally long — it lasted for at least half an hour — and the X-ray afterglow faded much faster than usual, an observation that is

incompatible with prevailing GRB models. The energy spectrum of the afterglow displayed a shape (the Planckian form) that is characteristic of a plasma whose electrons arrive at such an equilibrium that we can say that it has a certain temperature. Conversely, typical energy spectra of GRBs and their afterglows follow power laws that well describe the collision of materials at high velocities, such as that resulting from the impact of an explosion on its environment.

If the Christmas GRB originates from a remote hypernova, as do long-lasting GRBs, why is it so unusual? Could it be a completely different phenomenon, only apparently similar to a GRB? So far, observations have not unambiguously disclosed a host galaxy. The location of the GRB in the sky is such that it could belong to the Perseus arm of the Milky Way, but also to a satellite of the Andromeda galaxy. In their studies, Campana *et al.*² and Thöne *et al.*³ set out to address these questions and to explain the origin of the GRB.

Thöne *et al.*³ interpret a change in slope and colour of the Christmas GRB afterglow's optical emission as being due to the addition of a supernova that emerges after 10 days, when the afterglow has almost vanished. By comparing the system with similar composite supernova-GRB objects, they derive a distance for the system of 1.6 gigaparsecs and an energy of about 1.4×10^{51} erg, compatible with a typical GRB.

Thöne and colleagues posit that the GRB originates from a tightly bound binary system composed of a neutron star and a supergiant helium star. The helium star would transfer mass to the neutron star, and the system would evolve to a phase in which the stars would be surrounded by a common gas envelope produced by the expansion of the external layers of the helium star. The neutron star and the core of the helium star would merge to form a black hole or a highly magnetized neutron star, known as a magnetar⁴, producing a GRB-like jet that would eventually emerge through the common envelope. The authors find that this model agrees with the observed thermal spectrum of the GRB. However, the case for the additional supernova, which would be unusually faint, is not compelling. Therefore, the deduced GRB distance from Earth is uncertain. Moreover, it isn't clear whether a host galaxy exists. The only optical object seen near the position at which the GRB took place is point-like and extremely faint: were it a galaxy it would be the faintest ever to host a GRB.

Meanwhile, Campana *et al.*² revive a model proposed⁵ in 1973, soon after the discovery of GRBs. In this model, a minor body, such as a comet or an asteroid, flying by a neutron star at a distance of less than 5,000 kilometres, gets tidally disrupted, breaks into fragments and produces a multi-peak GRB. The authors² used recent models for tidal-disruption events⁶ and further assumed that the debris of the minor body forms a transient disk around the star. They compared these models with the available

data for the Christmas GRB and found that a model in which a minor body of 5×10^{20} grams falls onto a neutron star in our Galaxy is in excellent agreement, in both temporal and spectral behaviour, with the data.

Both hypotheses^{2,3} are plausible and explain numerous and complex data. But at least one is wrong and the definitive proof — namely, unambiguous determination of the GRB's distance from Earth — is missing. Thöne and colleagues' study raises several questions. How many binaries pass through a phase of a common envelope and produce GRB-like explosions? Do all jets emerge from the envelope, or is that the case for only a few of them? How highly collimated is the jet? But Campana and colleagues' study also opens up queries. How many minor bodies pass in close proximity to neutron stars? And is the deduced mass for the minor body conceivable? The proposed mass is relatively large (and consequently rare) for an object in the Solar System, but larger bodies have actually been found to orbit neutron stars. We could

compute the likelihood of each hypothesis, and perhaps discard one on the basis of statistical considerations. Unfortunately, however, such a computation would inevitably involve considerable conjecture.

In short, not much more can be said about the nature of the Christmas GRB — except that the odds are that the event is a rare phenomenon that looks like a GRB but falls outside this category. Whatever the case, it's hard to escape the fascination of a possible comet death on Christmas Day. ■

Enrico Costa is at IASF-Roma, INAF, Via del Fosso del Cavaliere 100, 00133 Roma, Italy. e-mail: enrico.costa@iasf-roma.inaf.it

1. Costa, E. *et al. Nature* **387**, 783–785 (1997).
2. Campana, S. *et al. Nature* **480**, 69–71 (2011).
3. Thöne, C. C. *et al. Nature* **480**, 72–74 (2011).
4. Barkov, M. V. & Komissarov, S. S. *Mon. Not. R. Astron. Soc.* **415**, 944–958 (2011).
5. Harwit, M. & Salpeter, E. E. *Astrophys. J.* **186**, L37–L39 (1973).
6. Lodato, G. & Rossi, E. M. *Mon. Not. R. Astron. Soc.* **410**, 359–367 (2011).

EARTH SCIENCE

Redox state of early magmas

A study of cerium in zircon minerals has allowed an assessment of the redox conditions that prevailed when Earth's earliest magmas formed. The results suggest that the mantle became oxidized sooner than had been thought. SEE LETTER P.79

BRUNO SCAILLET & FABRICE GAILLARD

A prime goal of petrologists has been to assess the redox conditions of magmas produced throughout Earth's evolution, because magmas are known to affect various major phenomena — such as the composition of volcanic emanations, which are widely believed to affect the composition of the atmosphere¹. But research efforts have been hampered by the lack of rocks from the Hadean eon, which encompasses nearly the first half-billion years of Earth's existence. On page 79 of this issue, Trail *et al.*² report their analysis of the sole mineral survivors of the Hadean, zircon samples more than 4 billion years old. Their findings allowed them to determine the 'fugacity' of oxygen in Hadean magmatic melts, a quantity that acts as a measure of magmatic redox conditions. Unexpectedly, the zircons record oxygen fugacities identical to those in the present-day mantle, leading the authors to conclude that Hadean volcanic gases were as highly oxidized as those emitted today.

The continuous reshaping of the surface

and deep interior of our dynamic planet has inevitably resulted in the loss of evidence of Earth's ancient composition. In this respect, Earth contrasts sharply with other planetary objects — particularly the Moon and Mars, which are largely unaffected by large-scale geodynamics. Thanks to the Apollo programme and to meteorites inferred to be of Martian origin, we have rock samples that provide direct testimonies of the Moon and Mars's very early past. But the same cannot be said of Earth. Nevertheless, it is generally assumed that all of these bodies were chemically reduced when they first formed and then underwent either a gradual or stepwise oxidation³.

One of the main arguments for a reduced start is that, during or soon after accretion, Earth must have entered a magma-ocean stage, when a sizeable part of the planet was molten because of the energy released by accretion processes³. A massive segregation process must then have occurred that separated metals from silicates, thus allowing the formation of the metallic core and the silicate-containing mantle. The formation of Earth's core is thought to have occurred 30 million to 60 million years

after the beginning of the Solar System. The coexistence of silicate and metal liquids in the magma ocean would have made conditions severely reduced, with oxygen fugacity (f_{O_2}) 5 to 8 log units below the current value for the upper mantle (Fig. 1).

Assuming that such a reduced starting point is inescapable, how did Earth reach its modern oxidized state, and how fast did this happen? Persistent oxidizing conditions at Earth's surface started when the planet was around 2.3 billion to 2.4 billion years old, a phenomenon known as the Great Oxidation Event¹ (GOE). Rock analyses⁴ suggest that the mantle's redox state has been similar to its current state from as early as 3.8 billion years ago. Trail *et al.*² wanted to look even further back in time, so they developed a technique for studying Hadean zircons, which are the only solids dating from the first 500 million years of Earth's history that are known to have survived unchanged to the present day. By analysing the cerium content of the zircons and developing a method to calibrate their results, the authors were able to determine f_{O_2} for Hadean magmas.

Their findings extend the mantle's oxidized realm to almost 4.4 billion years ago. Although somewhat tenuous, this is the first direct evidence of the redox state of the earliest Earth. If the zircons analysed by the authors are representative of the Hadean eon, this result shrinks the duration of the reduced era of Earth's mantle to less than 150 million years. It also increases the lag time between the oxidation of the mantle and the subsequent oxidation of the atmosphere (the GOE; Fig. 1), which might seem to make it difficult to establish a direct connection between the two events. However, redox changes upon the release of gas from magmas⁵ and changes in the pressure of volcanic degassing⁶ could explain why the atmospheric GOE happened so long after the mantle's oxidation. But what was the cause of the mantle's great oxidation event?

Planetary-scale geodynamics is often cited as a mechanism for bringing oxidized material into the deep, reduced mantle, and water is thought to be responsible for oxidation⁷. This concept is deeply entrenched among geologists, although the details of the redox mechanisms involved remain unknown. Although a matter of strong debate^{8,9}, Earth's current budget of volatile compounds (including water) seems to have been reached no later than 100 million years after core formation, and so volatiles may have fuelled the mantle's great oxidation at about that time (Fig. 1). However, it is known that terrestrial magmas do not necessarily have a high oxidation state, even if volatiles are abundant^{10,11}, and that the Moon's reduced lavas contain a substantial amount of volatiles¹². Taken together, these facts argue against the existence of an unequivocal relationship between high f_{O_2} and the abundance of volatiles.

But the accretion of volatiles into Earth's

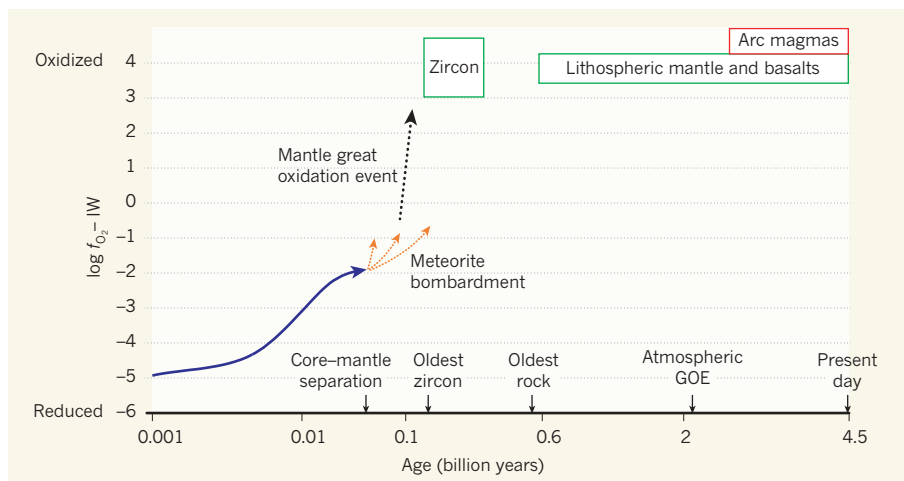


Figure 1 | Evolution of the redox state of Earth's mantle. The oxygen fugacity (f_{O_2} , measured in bar) of the mantle — a measure of the mantle's redox state — has varied over time. Here, $\log f_{O_2}$ is plotted relative to a standard value (known as the iron-wüstite buffer; IW); the scale on the x-axis is logarithmic. At first, the mantle was highly chemically reduced, but it became more oxidized as Earth's accretion proceeded and as core-mantle redox equilibration occurred (blue arrow). Meteorite bombardment may have contributed to oxidation of the mantle⁸ (orange dotted arrows) once it had segregated from the core. A 'great oxidation event' in the mantle (black dotted arrow) then occurred, in which f_{O_2} rapidly increased. On the basis of their analysis of 4.4-billion-year-old zircon samples, Trail *et al.*² report that the mantle's redox state about 100 million years after core-mantle separation (green box, left) was similar to that of the present-day lithospheric mantle (green box, right). The most highly oxidized mantle melts are the most recent — the arc magmas (red box) that formed as a result of crust-mantle exchanges triggered by subduction. The timing of some key events in Earth's history are indicated, along with the times when the oldest rock and zircon samples were formed; GOE is the Great Oxidation Event.

bulk is not the only possible mechanism for oxidation of the mantle. Experiments have shown¹³ that, at the pressure of the lower mantle, iron(II) oxide is converted into iron metal and iron(III) oxide — which means that large bodies such as Earth can self-oxidize their mantle, whereas smaller ones cannot (or do so to a lesser extent). In other words, a nominally dry mantle can be oxidized. This mechanism, however, would require that wholesale mantle convection and mixing occurred within a short period of time (less than 100 million years).

Alternatively, changes in the mantle's redox state could have occurred if the early magma ocean crystallized and degassed, because of the redox effect of degassing⁵ and the greater affinity of iron(III) for liquid relative to minerals¹⁴ (the iron(III) would have become concentrated in the residual liquid, thus increasing the oxidation state of the liquid). Such processes would probably have affected the entire mantle and would not have required large-scale convection.

Clearly, many parts of Earth's early evolution are still obscured by poor preservation of the geological record and by our limited knowledge of the mechanisms that drive magmatic redox patterns. Even so, the zircon record revealed by Trail *et al.*² firmly anchors one of the first redox steps of the infant Earth. This extraordinary tale is certainly not finished, and future work will augment the existing data — perhaps the discovery of other zircon samples will allow f_{O_2} to be determined for even more highly reducing conditions at

earlier periods of Earth's history, thereby refining our knowledge about what f_{O_2} represents. Also much needed are experimental or modelling studies aimed at understanding the mechanisms of redox processes in magma, and the extent to which redox state reflects either the origins of magmas or later processes that affect them. ■

Bruno Scaillet and Fabrice Gaillard are at the CNRS/INSU-Université d'Orléans, ISTO, UMR 6113, 1a rue de la Férollerie, 45071 Orléans, France.
e-mails: bscaille@cnrs-orleans.fr;
fabrice.gaillard@cnrs-orleans.fr

- Holland, H. D. *Geochim. Cosmochim. Acta* **66**, 3811–3826 (2002).
- Trail, D., Watson, E. B. & Tailby, N. D. *Nature* **480**, 79–82 (2011).
- Wood, B. J., Walter, M. J. & Wade, J. *Nature* **441**, 825–833 (2006).
- Li, Z.-X. A. & Lee, C.-T. A. *Earth Planet. Sci. Lett.* **228**, 483–493 (2004).
- Burgisser, A. & Scaillet, B. *Nature* **445**, 194–197 (2007).
- Gaillard, F., Scaillet, B. & Arndt, N. T. *Nature* **478**, 229–232 (2011).
- Kelley, K. A. & Cottrell, E. *Science* **325**, 605–607 (2009).
- Albarède, F. *Nature* **461**, 1227–1233 (2009).
- Wood, B. J., Halliday, A. N. & Rehkämper, M. *Nature* **467**, E6–E7 (2010).
- Mallmann, G. & O'Neill, H. St C. *J. Petrol.* **50**, 1765–1794 (2009).
- Scaillet, B. & Macdonald, R. J. *Petrol.* **42**, 825–845 (2001).
- Saai, A. E. *et al. Nature* **454**, 192–195 (2008).
- Frost, D. J., Mann, U., Asahara, Y. & Rubie, D. C. *Phil. Trans. R. Soc. A* **366**, 4315–4337 (2008).
- Carmichael, I. S. E. *Contrib. Mineral. Petrol.* **106**, 129–141 (1991).

The case against climate regulation via oceanic phytoplankton sulphur emissions

P. K. Quinn¹ & T. S. Bates¹

More than twenty years ago, a biological regulation of climate was proposed whereby emissions of dimethyl sulphide from oceanic phytoplankton resulted in the formation of aerosol particles that acted as cloud condensation nuclei in the marine boundary layer. In this hypothesis—referred to as CLAW—the increase in cloud condensation nuclei led to an increase in cloud albedo with the resulting changes in temperature and radiation initiating a climate feedback altering dimethyl sulphide emissions from phytoplankton. Over the past two decades, observations in the marine boundary layer, laboratory studies and modelling efforts have been conducted seeking evidence for the CLAW hypothesis. The results indicate that a dimethyl sulphide biological control over cloud condensation nuclei probably does not exist and that sources of these nuclei to the marine boundary layer and the response of clouds to changes in aerosol are much more complex than was recognized twenty years ago. These results indicate that it is time to retire the CLAW hypothesis.

Cloud condensation nuclei (CCN) can affect the amount of solar radiation reaching Earth's surface by altering cloud droplet number concentration and size and, as a result, cloud reflectivity or albedo¹. CCN are atmospheric particles that are sufficiently soluble and large enough in diameter to take up water vapour and serve as nuclei for cloud droplet formation. For the range of water supersaturations typical of marine boundary layer (MBL) clouds and the chemical composition of marine aerosols, the CCN population is dominated by particles between 40 and 300 nm in diameter. Although particles larger than 300 nm will activate to form cloud droplets regardless of composition, there are relatively few particles of this size. It is therefore the particles less than 300 nm in diameter that determine the CCN concentration in the remote MBL and that have the potential to change cloud properties.

Marine phytoplankton produce the osmolyte dimethylsulphonium propionate, which undergoes enzymatic cleavage to form dimethyl sulphide (DMS). Both dimethylsulphonium propionate and DMS scavenge reactive oxygen species and act as antioxidants under conditions of high ultraviolet radiation and oxidative stress^{2,3}. DMS is the dominant volatile sulphur compound in ocean surface waters and is the most significant biological source of gaseous sulphur to the remote marine troposphere⁴. After emission to the atmosphere, DMS is oxidized by the hydroxyl radical to form SO₂ and methane sulphonie acid (MSA)⁵. SO₂ can be taken up by particles directly or be further oxidized to H₂SO₄. H₂SO₄ can combine with other gas-phase species to form solid particles (homogeneous nucleation), while both H₂SO₄ and MSA can condense onto existing particles contributing to the growth of those particles (heterogeneous nucleation). The end result is a particulate phase containing SO₄²⁻ and, to a lesser extent, methane sulphonate (MSA⁻) derived from oceanic DMS.

In the 1980s, it was hypothesized that DMS-derived SO₄²⁻ made up the majority of the CCN in the MBL distant from continental and anthropogenic aerosol sources^{6,7}. The CLAW hypothesis, named after the four authors of the Charlson *et al.*⁷ study, was based on data available at the time that indicated that (1) non-sea-salt sulphate was ubiquitous in submicrometre marine aerosol^{8,9} and (2) concentrations of sodium containing particles at cloud height were negligible (thus ruling out sea salt particles as a source of CCN)¹⁰. Organic species were not considered because little was known about their concentration and composition in the marine atmosphere.

The CLAW hypothesis further postulated that an increase in DMS emissions from the ocean would result in an increase in CCN, cloud droplet concentrations, and cloud albedo, and a decrease in the amount of solar radiation reaching Earth's surface. A reduction in solar radiation could result in a change in the speciation and abundance of phytoplankton producing dimethylsulphonium propionate, thus setting up a climate feedback loop between cloud albedo and surface ocean DMS concentrations (Fig. 1). The proposed climate feedback loop requires (1) that DMS is a significant source of CCN to the MBL, (2) a change in DMS-derived CCN yields a change in cloud albedo, and (3) a change in cloud albedo, surface temperature, and/or incident solar radiation leads to a change in DMS production. If any one step in the feedback loop shown in Fig. 1 has a small response, the proposed bio-regulation of climate will be minimal.

The introduction of the CLAW hypothesis spawned 25 years of research in the fields of biological oceanography, atmospheric chemistry, and climate science in search of evidence for a biological regulation of climate due to marine sulphur emissions. Below we re-evaluate the individual steps in the feedback loop in the context of results from measurement and modelling studies conducted over the past two decades and conclude that a DMS bioregulation of climate is prevented by a weak sensitivity to change in each step of the feedback loop.

DMS as a source of CCN to the remote MBL

The strongest evidence for the first step in the CLAW feedback loop (the link between DMS emissions and CCN concentrations) is the coherence in the seasonal cycles of DMS, its particulate phase oxidation products (non-sea-salt SO₄²⁻ and MSA⁻), and CCN. At Cape Grim, Tasmania, a remote Southern Ocean sampling site, concentrations of atmospheric DMS, non-sea-salt SO₄²⁻, MSA⁻ and CCN all peak in the summer and are at a minimum in the winter as is expected for parameters controlled by biological activity^{11,12}. Similar measurements of DMS and CCN made in the boundary layer over the tropical South Atlantic¹³ and the north-east Pacific Ocean¹⁴ found that 40% to 50% of the variance in CCN concentrations could be explained by DMS. The South Atlantic observations also revealed significant correlations between non-sea-salt SO₄²⁻ or MSA⁻ mass concentrations and CCN¹³. Although these observations provide evidence for the atmospheric chemistry portion of the CLAW

¹NOAA Pacific Marine Environmental Laboratory, Seattle, Washington 98115, USA.

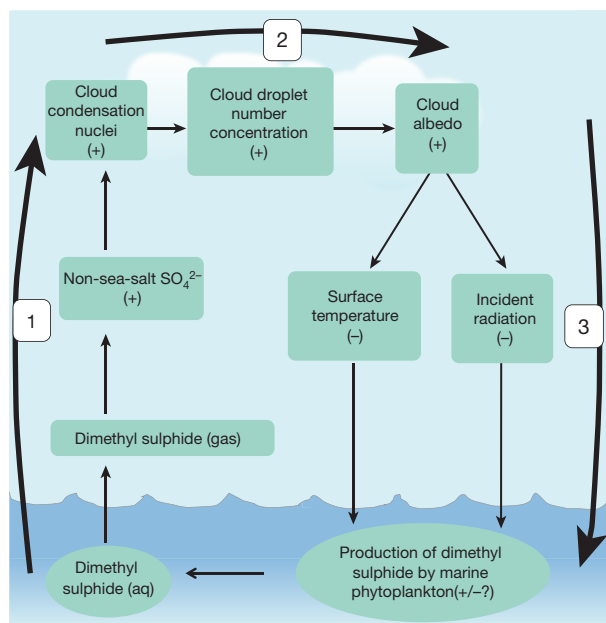


Figure 1 | Modified diagram of the climate feedback loop proposed by ref. 7. DMS, a breakdown product of phytoplankton-produced dimethylsulphonium propionate, undergoes sea-to-air transport. Once in the atmosphere, it is oxidized and these oxidation products undergo homogeneous nucleation to form new, water-soluble particles in the MBL. These particles become CCN and increase the number concentration of cloud droplets and cloud albedo. The resulting decrease in temperature at the surface of Earth and incident solar radiation then affect the rate of DMS production. The proposed climate feedback loop requires a strong response in each of the following steps: (1) a change in seawater DMS production leads to a significant change in MBL CCN concentration, (2) a change in DMS-derived CCN yields a change in cloud albedo, and (3) a change in cloud albedo, surface temperature, and/or incident solar radiation leads to a change in DMS production. The three steps are labelled in the figure.

hypothesis (that is, the conversion of SO_4^{2-} and MSA^- into CCN shown in step 1 of Fig. 1), they do not indicate the sensitivity of CCN concentration to a change in DMS emission. In addition, they do not rule out other sources of CCN. Measurements of the chemical composition of particles in the CCN size range are needed to assess all sources of CCN to the MBL.

Non-DMS sources of MBL CCN

Identification of the sources that contribute to MBL CCN requires direct chemical analysis of single particles with diameters less than 300 nm. Such measurements are sparse. Those measurements that have been reported are based on electron microscopy coupled with X-ray analysis^{15–19}, X-ray spectromicroscopy^{20,21} and laser mass spectrometry¹⁸. In addition, chemical composition has been inferred by comparing unheated and heated number size distributions with subsequent quantification of the volatile and refractory fractions of the aerosol²². Below we discuss what is known about the composition of CCN in the remote MBL on the basis of these measurement techniques and reveal the significance of sea salt and non-DMS organics.

Single-particle analysis of aerosol samples collected over the open ocean in the mid-Pacific reveal that non-sea-salt SO_4^{2-} exists in CCN-size particles^{15,16} and that the number concentration in this size range can be dominated by particles containing sulphate¹⁷. At the same time, there is a lack of observations of pure acidic SO_4^{2-} or $(\text{NH}_4)_2\text{SO}_4$ particles in the MBL. Measurements made on the outer edge of the Great Barrier Reef and over the central Arctic Ocean reveal organic inclusions within $(\text{NH}_4)_2\text{SO}_4$ particles¹⁹. The organic inclusions appear to act as nuclei

for the condensation of gas-phase DMS oxidation products. Similarly, measurements of aerosol composition at Cape Grim indicate that particles formed through the condensation of sulphur-containing gases grow only by coagulating with sea salt and organics¹⁸. DMS will increase the CCN number concentration through coagulation and condensation only if it allows a particle to grow and/or become sufficiently soluble to activate to form a cloud droplet when entrained into a cloud. If the sea salt or organic particle was already sufficiently large to serve as a CCN, the addition of DMS-derived sulphur to the particle will not increase the number of CCN.

Sea salt as a source of MBL CCN

Unambiguous identification of sea salt as CCN requires the determination of the number concentration of sea-salt-containing particles in those particles with diameters less than 300 nm. Though few, there are measurements based on direct chemical techniques that provide evidence for sea salt as CCN. Measurements on the coast of Oahu, Hawaii, with an instrument that detects emission from thermally excited sodium atoms reveal the presence of sodium-containing particles down to the instrumental cut-off diameter of 200 nm (ref. 23). Manual transmission electron microscopy with energy dispersive X-ray analysis of aerosol samples collected over the Southern Ocean finds that up to 50% of the particles in the 50–150-nm size range contain sea salt¹⁸. This percentage corresponds to a number concentration of sea-salt-containing particles of up to 100 per cubic centimetre, a significant fraction of the typical MBL CCN concentration²².

Indirect measurements based on volatility have been used over the past several decades to infer the contribution of sea salt to the remote MBL particle population. Volatile, or non-refractory, aerosol chemical species such as H_2SO_4 , ammonium sulphates, and simple organics evaporate or decompose at temperatures below 300 °C (ref. 24). Refractory species, including dust, sea salt and large organic polymers, must be heated to over 600 °C for thermal decomposition to occur^{24,25}. Flights across the Atlantic reveal that, at altitudes near 400 m, 20% of the measured CCN survived heating to 650 °C (ref. 25). The fraction of refractory nuclei decreased with altitude, indicating an ocean source. A similar volatility analysis onboard a research cruise in the northeast Atlantic found that for maritime, high-wind-speed air masses ($14\text{--}17\text{ m s}^{-1}$), 40% of the particles with diameters between 100 and 240 nm contained a refractory component like sea salt²⁶.

The most compelling evidence for sea salt as CCN is provided by single-particle chemical analysis of residual particles that remain after water is evaporated from cloud droplets. Measurements of aerosol particles that were sampled below stratocumulus clouds over the southeast Pacific revealed that half of the particles less than 200 nm in diameter contained sulphate. In contrast, the majority (60%) of the residual particles from evaporated cloud droplets—that is, the particles that had acted as CCN—was sea salt²⁷. This result is due to the lower supersaturation required for sea salt to activate and form cloud droplets compared to other species, including sulphate, for a given particle size. The experiments described above, as well as others carried out over the past several decades^{22,28}, show that sea-salt-containing particles in the CCN size range are present in marine air masses. Hence, contrary to what was assumed 25 years ago, observations show that sea salt makes up a large fraction of MBL CCN.

Organics as a source of MBL CCN

Ocean surface waters contain large concentrations of small particulates including phytoplankton, algae, bacteria, viruses, fragments of larger organisms and organic detritus^{29,30}. Phytoplankton release or exude organic matter during growth, predation by grazing organisms and viral lysis³¹. The resulting pool of organic matter contributes to one of the largest active reservoirs of organic carbon on Earth³² and consists of every class of carbohydrate³¹. This pool of organic matter is often referred to as dissolved organic matter but may contain colloidal material that is not truly dissolved. A growing body of evidence shows that this

seawater-dissolved organic matter is similar to organic material found in atmospheric aerosols^{33–35}.

Phytoplankton exudates include exopolymer gels consisting of polysaccharides that bind together smaller organic molecules such as amino acids, proteins and peptides³⁶. Exopolymer gels are insoluble, thermally stable, highly surface active, highly hydrated and readily sequester dissolved organic matter³⁷. Exopolymer gels bind small organic particulates into aggregates. It has been hypothesized that these aggregates are injected into the atmosphere through bubble bursting at the ocean surface, thereby contributing to the primary organic aerosol found in the remote MBL^{19,35}. Once in the atmosphere, the exopolymer gels that bind the aggregates and particles together are degraded with ultraviolet light or acidification. Transmission electron microscopy analysis on particles collected during the summer at Cape Grim and other remote ocean sites have revealed the existence of individual particles with diameters between 70 and 200 nm containing organic matter in the form of exopolymer gels and airborne marine aggregates composed of exopolymer gels, aggregates of organic particles and marine microorganisms^{19,35,37}. It has been suggested that the size and solubility continuum of phytoplankton exudates that exists in sea water is reflected in nascent sea spray particles³³. This mechanism for the production of organic sea spray aerosol is depicted in Fig. 2.

Observational evidence for an organic source of CCN to the MBL is considerable. Measurements of northeast-Atlantic air masses made on the coast of Ireland reveal increased mass concentrations of organics, primarily water insoluble, in sub-500-nm particles during the summer bloom period³⁸. Aerosol organic matter recently measured in samples collected in remote regions of the North Atlantic and Arctic oceans is reported to be composed of carbohydrate-like compounds containing organic hydroxyl groups from ocean emissions²⁰. The abundance of carbohydrate material measured in these aerosol samples is consistent with observations made over the past decade. A review of over ten

years of measurements of marine organic aerosol concluded that this ocean-derived aerosol is composed of polysaccharides, proteins and amino acids, and microorganisms and their fragments²¹.

Recently, several studies have been performed involving the generation of nascent aerosols from sea water to avoid interference from constituents in the marine atmosphere that could modify newly emitted particles³⁹. Despite the use of different methods in these experiments, several generalities can be made regarding the characteristics of nascent, ocean-derived particles. All experiments show a dominant submicrometre mode in the particle number size distribution with a peak diameter between 50 and 100 nm (refs 40–42). In addition, organics are significantly enriched in submicrometre particles with the degree of enrichment increasing with decreasing particle size^{33,39}. These studies, along with the similarities between the composition of organic matter sampled in the ocean microlayer and in atmospheric aerosol⁴³, provide further evidence for an organic, wind-driven source of CCN to the MBL.

The MBL CCN budget

Under the CLAW hypothesis, an increase in the emission of DMS from the ocean results in an increase in the number of particles that act as CCN by either (1) the homogeneous nucleation of new particles in the MBL involving H_2SO_4 or (2) the growth of existing particles to CCN-size through condensation of SO_2 , H_2SO_4 or MSA^- . Whether gas-phase species undergo homogeneous nucleation or condense onto existing particles depends, in part, on the surface area of the existing particles. If sufficient surface area is available, it provides a sink for condensable vapours and new particle formation is prevented^{44,45}. Model calculations indicate that if precipitation removes pre-existing aerosol, thereby reducing aerosol surface area, homogeneous nucleation of new particles can occur under moderate DMS concentrations⁴⁶. There is observational evidence for particle nucleation in the MBL under conditions of low particle surface area^{44,47}, but it is limited. In addition, in these and other cases^{48–50}, there is no evidence that the events led to particles large enough to be CCN. Hence, for conditions most typical of the marine atmosphere (large surface area of existing, primarily sea-salt particles and low SO_2 concentrations⁵¹), boundary layer nucleation due to DMS-derived sulphur and water vapour has little theoretical⁴⁶ or observational basis.

Unlike the remote MBL, observations of H_2SO_4 -initiated particle nucleation in the free troposphere near cloud top height are numerous^{47,52–55}. Gases, including DMS, and particles are entrained from the MBL into clouds. Clouds scavenge or capture the particles so that the air that is detrained from the cloud contains very low aerosol surface area concentrations⁵³. The nucleation of new particles observed at cloud top height is a result of reactive gases such as DMS being pumped up to the upper troposphere through deep convective clouds. Once aloft, the reactive gases undergo gas-to-particle conversion in cloud outflow regions where the existing particle surface area is low, water vapour concentrations are high, and the actinic flux is high⁴⁷. Simple box models⁵⁶ as well as advanced three-dimensional global models^{57–60} indicate that subsidence of particles from the free troposphere can explain a significant fraction of the CCN measured in the MBL. Hence, measurements and model calculations published since the introduction of the CLAW hypothesis indicate that DMS-derived sulphate contributes to MBL CCN concentrations via particle nucleation in the free troposphere rather than in the boundary layer itself. After formation in the free troposphere, the particles may be transported thousands of kilometres before mixing down into the MBL. As a result, regions of high-DMS sea-to-air fluxes do not correlate with regions of high DMS-derived CCN concentrations⁶¹. The spatial decoupling between DMS production and entrainment of DMS-derived particles into the MBL prevents a local marine biota–climate feedback loop^{60–62}.

DMS emissions are required for models to reproduce the observed seasonal cycle in CCN at Cape Grim. Over the Southern Ocean (30° S to 45° S) model estimates indicate that DMS emissions increase CCN concentrations during the summer by 46% with emissions of sea salt

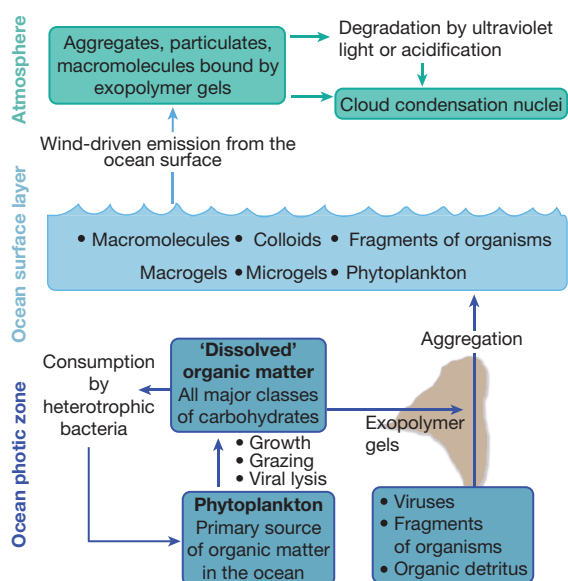


Figure 2 | Ocean-derived source of organics to the MBL CCN population.

Phytoplankton are the primary source of organic matter in the ocean. The large pool of dissolved organic matter composed of all major classes of carbohydrates resides within the ocean photic zone and is derived from phytoplankton growth, grazing by predators and viral lysis. Also residing within the photic zone are small particulates composed of other algae, bacteria, viruses, fragments of larger organisms and organic detritus. This organic matter becomes bound together by exopolymer gels secreted by marine organisms. The resulting aggregates are injected into the atmosphere through bubble bursting and contribute to the MBL CCN population either directly or after degradation by ultraviolet light or acidification.

accounting for the majority of the remainder⁶¹. The majority of the summertime DMS-derived CCN is a result of particle nucleation in the free troposphere. Only a small fraction (6%) of DMS oxidation products contribute to condensational growth of sea salt to CCN sizes because most of the sea salt is large enough to be CCN without further growth. Further south ($>45^\circ$ S), the contribution of DMS to the MBL CCN population decreases owing to high wind speeds that result in a large sea-salt source. During winter, sea-salt emissions account for over 80% of CCN. Additional modelling results confirm a significant sea-salt source of CCN. In remote ocean regions under conditions of high wind speeds, emissions of sea salt have been calculated to increase CCN concentrations up to 500% above a DMS-derived CCN background concentration⁵⁷. The inclusion of emissions of sea-salt particles with diameters between 10 and 100 nm is required to explain observed particle number concentrations in these regions. Finally, model results indicate that ocean emissions of organics provide a significant source of sub-micrometre aerosol mass and number to the MBL⁵⁹. Without inclusion of organics, CCN concentrations in the MBL are underestimated⁶³. The large contribution of wind-driven sea spray containing both sea salt and organics to the MBL CCN population prevents DMS from being the sole source of MBL CCN, which weakens step 1 of the CLAW feedback loop.

Sea spray aerosol production fluxes

Although recent advancements in global models have allowed for the estimation of MBL CCN budgets, large uncertainties in these estimates remain due, in large part, to uncertainties in sea-spray aerosol production fluxes. Model calculations of the diameter and number concentration of sea-spray-derived CCN are highly sensitive to emission parameterizations of sea salt and organics. Yet estimates of the size-dependent production flux of sea-salt aerosol vary by an order of magnitude, with even larger differences occurring in the size range most relevant for CCN formation⁶⁴. Furthermore, accurate parameterizations of oceanic emissions of organics are complicated by the patchiness of biological productivity in ocean surface waters. Amounts of chlorophyll *a* derived from satellite observations have been used as a proxy for ocean biological activity in efforts to parameterize emissions of organics⁶⁵. This approach is only able to account for 25% of the variance in the water insoluble organic aerosol mass concentration, indicating that chlorophyll represents only a small fraction of the parameters involved in controlling concentrations of organics in ocean surface waters.

The large pool of dissolved organic matter resulting from phytoplankton secretions has not been detected with satellite retrievals of chlorophyll, nor have non-phytoplankton particulates (bacteria, viruses, fragments of larger organisms and organic detritus). Hence, there is a need for a sea-spray source function that accounts for this additional source of organic matter. Recently developed source functions of sea-spray-derived organic aerosol include more complex biological processes whereby the organic component is derived from seawater diatoms⁴². The parameterization incorporates retrievals of chlorophyll *a* diatom biomass from satellites and first-order approximations of the amount of total organic matter resulting from diatom growth, predation by grazing organisms and viral lysis. Although large uncertainties in the source function exist, this approach acknowledges the complexities involved in relating seawater organic matter to emissions of sea spray organic aerosols. Field observations involving multiple experimental approaches are required to constrain the production fluxes of sea spray (including sea salt and organics) and to reduce the uncertainties in model-calculated distributions of CCN over the world's oceans⁶⁴.

The DMS–cloud albedo climate feedback

Global-scale models have been used to determine the sensitivity of CCN to an increase in the emission of DMS for both present-day and globally warmed scenarios⁶⁶. For the Southern Hemisphere, the summertime sensitivity for present-day DMS fluxes is about 0.07, indicating a 0.07% increase in total CCN (from all sources) that is attributable to a 1% increase in the flux of DMS. The sensitivity is lower for the Northern

Hemisphere (0.02%), where the aerosol abundance from all sources is higher. Global warming is predicted to lead to a shallowing of the ocean mixed-layer depth accompanied by an increase in solar radiation dose in the upper mixed layer. Under a globally warmed scenario of a 50% increase in CO_2 , one model estimate indicates that the DMS flux will increase by 1%, which yields only a $\sim 0.1\%$ increase in the global CCN concentration⁶⁶. This low sensitivity is due to the abundance of MBL CCN derived from non-DMS sources. Other coupled ocean–atmosphere general circulation models have been used to study the impact of climate change on marine DMS emissions and CCN and have found only a weak positive global response of sulphate aerosol to increased CO_2 (ref. 67). As a result, only very large changes in the flux of DMS to the atmosphere would result in a significant impact on CCN concentration. This low sensitivity between DMS flux and CCN concentration contributes to the weak link in step 1 of the CLAW hypothesis.

The CLAW hypothesis (as proposed) assumed that, for a fixed content of liquid water in cloud, an increase in the concentration of CCN would lead to an increase in cloud albedo (see step 2 of Fig. 1). Charlson *et al.*⁷ calculated that a 30% increase in CCN concentration would cause an increase in planetary albedo sufficient to decrease global mean surface temperature by 1.3 K. As described above, the sensitivity of CCN to the sea-to-air flux of DMS is very low, such that a 30% increase in CCN concentration would require a 300% increase in the DMS flux. Furthermore, studies performed over the past decade with high-resolution models have revealed mechanisms that occur on cloud micro- and macro-physical scales that can buffer or are counter to an increase in cloud albedo that would be attributable to an increase in CCN^{68,69}.

Aerosols can affect not only cloud microphysics (cloud droplet size and number concentration) but also cloud macrophysics (cloud fraction, size and morphology)^{70,71}. For example, for a regime of non-precipitating clouds, dynamic feedbacks resulting from aerosol perturbations on clouds can decrease cloud fraction⁷⁰. An increase in CCN concentration results in faster evaporation rates owing to smaller cloud droplets. The faster evaporation rate leads to enhanced entrainment of subsaturated air surrounding the cloud and a decrease in cloud fraction. In this case, the aerosol effect of an increase in cloud albedo is lessened by the reduction in cloud fraction⁷¹. Parameterizations that do not consider aerosol-induced changes in cloud macrophysics may misrepresent the overall impact of aerosol–cloud interactions. By assuming a high sensitivity of CCN to the flux of DMS and considering only a highly simplified relationship between aerosol and cloud albedo, Charlson *et al.*⁷ probably significantly overestimated the response of cloud albedo to changing CCN concentrations in step 2 of the CLAW hypothesis.

Step 3 of the CLAW hypothesis assumes that changes in cloud albedo, surface temperature and/or incident solar radiation lead to changes in the production of DMS. This step is supported by time-series measurements that show a seasonal correlation between levels of downwelling solar radiation and ocean surface mixed-layer DMS concentrations^{3,72}. In the low-chlorophyll waters near Bermuda, upper-ocean DMS concentrations appear to be related to the availability of incident ultraviolet radiation⁷². This correlation does not reveal information about the sensitivity of DMS production to changes in solar radiation, however. Model calculations indicate that an increase in the solar radiation dose to the ocean upper mixed layer due to a 50% increase in the atmospheric concentration of CO_2 results in only a $\sim 1\%$ increase in global DMS ocean surface concentrations⁷³. Other model simulations indicate that changes in DMS emissions due to changes in temperature and irradiance that affect ocean primary productivity are small⁷⁴. This low sensitivity of seawater DMS concentrations to a global warming scenario indicates that the response in step 3 of the CLAW feedback loop is weak.

The post-CLAW view of MBL CCN

In 1987, Charlson *et al.*⁷ hypothesized that oxidation products of biogenically produced DMS account for the bulk of the CCN number concentration in the remote MBL, thus providing a link between marine biota and climate. Field and laboratory experiments performed over the

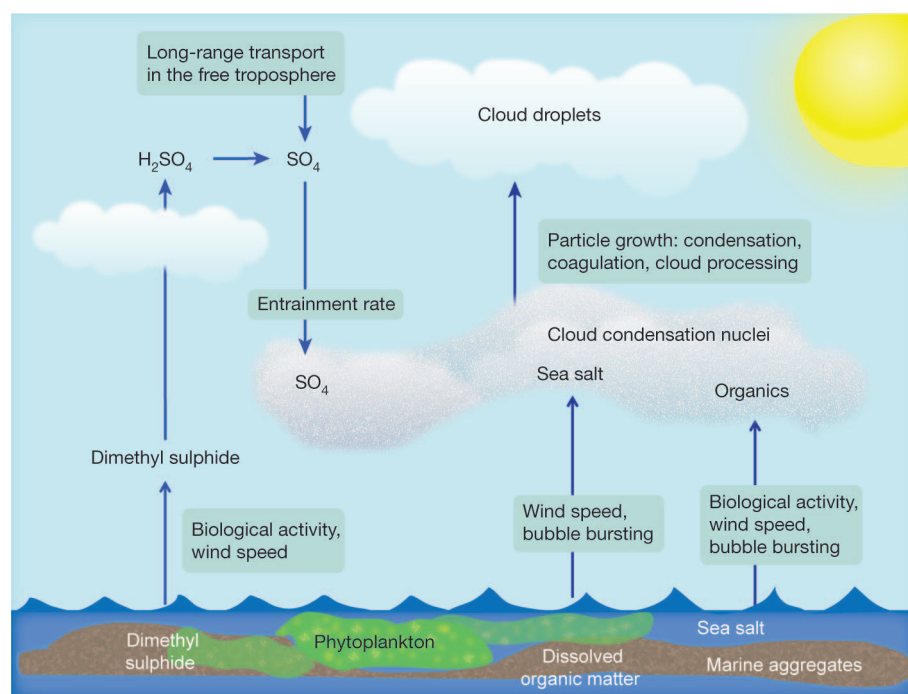


Figure 3 | Major sources and production mechanisms for CCN in the remote MBL. DMS contributes to the MBL CCN population primarily via particle nucleation in the free troposphere in cloud outflow regions with subsequent subsidence. Sea salt and organics are emitted as a result of wind-driven bubble bursting.

past decade have shown that sources of CCN to the remote MBL are more complex than was recognized by the CLAW hypothesis. Bubble bursting at the ocean surface is a major source of not just aerosol mass but also aerosol number to the MBL. This process introduces both inorganic and organic components of sea water to the atmosphere. Inorganic components are comprised of sea salt while the organic components are derived from phytoplankton and the large pool of organics in the ocean surface. Hence, the concentration of CCN in the remote MBL is a result of emissions of sea salt and organics in sea spray (dependent upon biological activity and wind speed), subsidence of DMS-derived and continentally derived particulates from the free troposphere (dependent upon oxidation and entrainment rates), and particle growth (dependent upon condensation, coagulation and cloud processing). This updated view of the multiple sources of CCN to the MBL is depicted in Fig. 3.

The evidence gained over the past 20 years of the significance of non-DMS sources of MBL CCN, the lack of observational evidence for a DMS-controlled marine biota–climate feedback, and the modelled low sensitivity between change and response in each step of the CLAW hypothesis feedback loop all indicate that it is time to retire the CLAW hypothesis. Retiring CLAW does not rule out a link between ocean-derived CCN and climate, however. Sea-salt aerosol production is expected to be affected by an increase in surface wind speed^{75,76} such as has been observed in association with spring Antarctic ozone depletion⁷⁷. In addition, because sea-spray organics are emitted as particulates directly into the atmosphere, a direct link between biological production and local emission of organics, CCN production and cloud albedo is possible.

When proposed⁷, the CLAW hypothesis, with its synthesis of marine biochemistry, atmospheric chemistry, cloud physics and climate dynamics into a plausible feedback loop, was visionary in scope. If the CLAW hypothesis has not stood the test of time, it is only because we now have a much better appreciation of the complexity of biogeochemistry and climate physics than when the hypothesis was first put forward. The interdisciplinary research that it motivated is now needed to address the complexity of multiple sources of CCN to the MBL and potential impacts on climate.

Received 14 January; accepted 21 September 2011.

- Twomey, S. The influence of pollution on the shortwave albedo of clouds. *J. Atmos. Sci.* **34**, 1149–1152 (1977).

- Sunda, W., Kieber, D. J., Kiene, R. P. & Huntsman, S. An antioxidant function for DMSP and DMS in marine algae. *Nature* **418**, 317–320 (2002).
- Vallina, S. M. & Simo, R. Strong relationship between DMS and the solar radiation dose over the global surface ocean. *Science* **315**, 506–508 (2007).
- Bates, T. S., Lamb, B. K., Guenther, A. B., Dignon, J. & Stoiber, R. E. Sulfur emissions to the atmosphere from natural sources. *J. Atmos. Chem.* **14**, 315–337 (1992).
- Andreae, M. O. *et al.* Dimethylsulfide in the marine atmosphere. *J. Geophys. Res.* **90**, 12891–12900 (1985).
- Shaw, G. E. Bio-controlled thermostasis involving the sulphur cycle. *Clim. Change* **5**, 297–303 (1983).
- Charlson, R. J., Lovelock, J. E., Andreae, M. O. & Warren, S. G. Oceanic phytoplankton, atmospheric sulphur, cloud albedo, and climate. *Nature* **326**, 655–661 (1987).
- This paper introduced the CLAW hypothesis proposing the link between marine biota and climate.**
- Andreae, M. O. Marine aerosol chemistry at Cape Grim, Tasmania and Townsville, Queensland. *J. Geophys. Res.* **87**, 8875–8885 (1982).
- Savoie, D. L. & Prospero, J. M. Particle size distribution of nitrate and sulphate in the marine atmosphere. *Geophys. Res. Lett.* **9**, 1207–1210 (1982).
- Hobbs, P. V. Simultaneous airborne measurements of cloud condensation nuclei and sodium-containing particles over the ocean. *Q. J. R. Meteorol. Soc.* **97**, 263–271 (1971).
- Ayers, G. P. & Gras, J. L. Seasonal relationship between cloud condensation nuclei and aerosol methanesulphonate in marine air. *Nature* **353**, 834–835 (1991).
- This paper presented the coherence in the seasonality of DMS-derived particulate phase sulphur species and CCN at Cape Grim, Tasmania.**
- Ayers, G. P., Caine, J. M., Gillett, R. W. & Ivey, J. P. Atmospheric sulphur and cloud condensation nuclei in marine air in the southern hemisphere. *Phil. Trans. R. Soc. Lond. B* **352**, 203–211 (1997).
- Andreae, M. O., Elbert, W. & de Mora, S. J. Biogenic sulphur emissions and aerosols over the tropical South Atlantic. 3. Atmospheric dimethylsulfide, aerosols, and cloud condensation nuclei. *J. Geophys. Res.* **100** (D6), 11335–11356 (1995).
- Hegg, D. A., Ferek, R. J., Hobbs, P. V. & Radke, L. F. Dimethyl sulfide and cloud condensation nucleus correlations in the northeast Pacific Ocean. *J. Geophys. Res.* **96** (D7), 13189–13191 (1991).
- Parungo, F. P., Nagamoto, C. T., Rosinski, J. & Haagenson, P. L. A study of marine aerosols over the Pacific Ocean. *J. Atmos. Chem.* **4**, 199–226 (1986).
- Pósfai, M., Anderson, J. R. & Buseck, P. R. Constituents of a remote Pacific marine aerosol: A TEM study. *Atmos. Environ.* **28**, 1747–1756 (1994).
- McInnes, L., Covert, D. & Baker, B. The number of sea-salt, sulfate, and carbonaceous particles in the marine atmosphere: EM measurements consistent with the ambient size distribution. *Tellus* **49B**, 300–313 (1997).
- Murphy, D. M. *et al.* Influence of sea-salt on aerosol radiative properties in the Southern Ocean marine boundary layer. *Nature* **392**, 62–65 (1998).
- This paper provided direct observational evidence of significant numbers of CCN-size particles containing sea salt and organics in the remote MBL.**
- Leck, C. & Bigg, E. K. Comparison of sources and nature of the tropical aerosol with the summer high Arctic aerosol. *Tellus* **60B**, 118–126 (2008).
- Russell, L. M., Hawkins, L. N., Frossard, A. A., Quinn, P. K. & Bates, T. S. Carbohydrate-like composition of submicron atmospheric particles and their production from ocean bubble bursting. *Proc. Natl Acad. Sci. USA* **107**, 6652–6657 (2010).

21. Hawkins, L. N. & Russell, L. M. Polysaccharides, proteins, and phytoplankton fragments: four chemically distinct types of marine primary organic aerosol classified by single particle spectromicroscopy. *Adv. Meteorol.* **2010**, 612132 (2010).
22. Clarke, A. D., Owens, S. R. & Zhou, J. An ultrafine sea-salt flux from breaking waves: implications for cloud condensation nuclei in the remote marine atmosphere. *J. Geophys. Res.* **111**, D06202 (2006).
23. Campuzano-Jost, P. *et al.* Near-real-time measurement of sea-salt aerosol during the SEAS campaign: comparison of emission-based sodium detection with an aerosol volatility technique. *J. Atmos. Ocean. Technol.* **20**, 1421–1430 (2003).
24. O'Dowd, C. D. & Smith, M. H. Physicochemical properties of aerosols over the northeast Atlantic: evidence for wind-speed-related submicron sea-salt aerosol production. *J. Geophys. Res.* **98** (D1), 1137–1149 (1993).
25. Dinger, J. E., Howell, H. B. & Wojciechowski, T. A. On the source and composition of cloud nuclei in a subsident air mass over the north Atlantic. *J. Atmos. Sci.* **27**, 791–797 (1970).
26. O'Dowd, C. D., Smith, M. H. & Jennings, S. G. Submicron particle, radon, and soot carbon characteristics over the Northeast Atlantic. *J. Geophys. Res.* **98**, 1123–1135 (1993).
27. Twohy, C. H. & Anderson, J. R. Droplet nuclei in non-precipitating clouds: composition and size matter. *Environ. Res. Lett.* **3**, 045002, 1–9 (2008).
This paper provided direct observational evidence of cloud droplets formed primarily through nucleation on sea-salt particles.
28. Peter, J. A., Blyth, A. M., Brooks, B., Lingard, J. & Smith, M. H. On the composition of Caribbean maritime aerosol particles measured during RICO. *Q. J. R. Meteorol. Soc.* **134**, 1059–1063 (2008).
29. Fuhrman, J. A. Marine viruses and their biogeochemical and ecological effects. *Nature* **399**, 541–548 (1999).
30. Wells, M. L. & Goldberg, E. D. Occurrence of small colloids in sea water. *Nature* **353**, 342–344 (1991).
31. Biersmith, A. & Benner, R. Carbohydrates in phytoplankton and freshly produced dissolved organic matter. *Mar. Chem.* **63**, 131–144 (1998).
32. Hedges, J. I. Global biogeochemical cycles: progress and problems. *Mar. Chem.* **39**, 67–93 (1992).
33. Facchini, M. C. *et al.* Primary submicron marine aerosol dominated by insoluble organic colloids and aggregates. *Geophys. Res. Lett.* **35**, L17814 (2008).
34. Leck, C. & Bigg, E. K. Evolution of the marine aerosol—a new perspective. *Geophys. Res. Lett.* **32**, L19803 (2005).
35. Bigg, E. K. Sources, nature, and influence on climate of marine airborne particulates. *Environ. Chem.* **4**, 155–161 (2007).
This paper described an organic alternative to DMS as a source of CCN to the MBL.
36. Decho, A. W. Microbial exopolymer secretions in ocean environments: their role(s) in food webs and marine processes. *Oceanogr. Mar. Biol. Ann. Rev.* **28**, 73–153 (1990).
37. Bigg, E. K. & Leck, C. The composition of fragments of bubbles bursting at the ocean surface. *J. Geophys. Res.* **113**, D11209 (2008).
38. O'Dowd, C. D. *et al.* Biogenically-driven organic contribution to marine aerosol. *Nature* **431**, 676–680 (2004).
39. Keene, W. C. *et al.* Chemical and physical characteristics of nascent aerosols produced by bursting bubbles at a model air-sea interface. *J. Geophys. Res.* **112**, D21202 (2007).
40. Tyree, C. A., Hellion, V. M., Alexandrova, O. A. & Allen, J. O. Foam droplets generated from natural and artificial seawaters. *J. Geophys. Res.* **112**, D12204 (2007).
41. Hultin, K. A. H. *et al.* In situ laboratory sea spray production during the Marine Aerosol Production 2006 cruise on the northeastern Atlantic Ocean. *J. Geophys. Res.* **115**, D06201 (2010).
42. Fuentes, E., Coe, H., Green, D., de Leeuw, G. & McFiggans, G. On the impacts of phytoplankton-derived organic matter on the properties of marine aerosol—Part 1: Source fluxes. *Atmos. Chem. Phys.* **10**, 9295–9317 (2010).
43. Bigg, E. K., Leck, C. & Tranvik, L. Particulates of the surface microlayer of open water in the central Arctic Ocean in summer. *Mar. Chem.* **91**, 131–141 (2004).
44. Covert, D. S., Kapustin, V. N., Quinn, P. K. & Bates, T. S. New particle formation in the marine boundary layer. *J. Geophys. Res.* **97**, 20581–20589 (1992).
45. Warren, D. R. & Seinfeld, J. H. Prediction of aerosol concentration resulting from a burst of nucleation. *J. Colloid Interf. Sci.* **105**, 136–142 (1985).
46. Pirjola, L., O'Dowd, C. D., Brooks, I. M. & Kulmala, M. Can new particle formation occur in the clean marine boundary layer? *J. Geophys. Res.* **105** (D21), 26,531–26,546 (2000).
47. Clarke, A. D. *et al.* Particle production in the remote marine atmosphere: cloud outflow and subsidence during ACE-1. *J. Geophys. Res.* **103** (D13), 16,397–16,409 (1998).
This paper was one of the first to provide unambiguous evidence of an upper tropospheric source of sulphur particles to the marine boundary layer.
48. Ehn, M. *et al.* in *Nucleation and Atmospheric Aerosols: 17th International Conference (Galway, Ireland, 2007)* (eds O'Dowd, C. D. & Wagner, P. E.) 1, 102–1, 105 (Springer, 2007).
49. Davison, B. *et al.* Dimethyl sulfide, methyl sulfonic acid, and physicochemical aerosol properties in Atlantic air from the United Kingdom to Halley Bay. *J. Geophys. Res.* **101**, 22,855–22,867 (1996).
50. O'Dowd, C. D. *et al.* Biogenic sulphur emissions and inferred non-sea-salt sulfate cloud condensation nuclei in and around Antarctica. *J. Geophys. Res.* **102**, 12839–12854 (1997).
51. Caine, J. & Harvey, M. Dimethylsulfide, a limited contributor to new particle formation in the clean marine boundary layer. *Geophys. Res. Lett.* **29**, 1128, doi:10.1029/2001GL014439 (2002).
52. Hegg, D. A., Radke, L. F. & Hobbs, P. V. Particle production associated with marine clouds. *J. Geophys. Res.* **95**, 13,917–13,926 (1990).
53. Perry, K. D. & Hobbs, P. V. Further evidence for particle nucleation in clear air adjacent to marine cumulus clouds. *J. Geophys. Res.* **99**, 22,803–22,818 (1994).
54. Hoppel, W. A., Frick, G. M., Fitzgerald, J. & Larson, R. E. Marine boundary layer measurements of new particle formation and the effects nonprecipitating clouds have on aerosol size distributions. *J. Geophys. Res.* **99**, 14,443–14,459 (1994).
55. Clarke, A. D., Li, Z. & Litchy, M. Aerosol dynamics in the equatorial Pacific marine boundary layer: microphysics, diurnal cycles, and entrainment. *Geophys. Res. Lett.* **23**, 733–736 (1996).
56. Raes, F. Entrainment of free-tropospheric aerosol as a regulating mechanism for cloud condensation nuclei in the remote marine boundary layer. *J. Geophys. Res.* **100**, 2893–2903 (1995).
57. Pierce, J. R. & Adams, P. J. Global evaluation of CCN formation by direct emission of sea salt and growth of ultrafine sea salt. *J. Geophys. Res.* **111**, D06203 (2006).
58. Kazil, J., Lovejoy, E. R., Barth, M. C. & O'Brien, K. Aerosol nucleation over oceans and the role of galactic cosmic rays. *Atmos. Chem. Phys.* **6**, 4905–4924 (2006).
59. Spracklen, D. V. *et al.* Evaluation of a global aerosol microphysics model against size-resolved particle statistics in the marine atmosphere. *Atmos. Chem. Phys.* **7**, 2073–2090 (2007).
60. Merikanto, J., Spracklen, D. V., Mann, G. W., Pickering, S. J. & Carslaw, K. S. Impact of nucleation on global CCN. *Atmos. Chem. Phys.* **9**, 8601–8616 (2009).
61. Korhonen, H., Carslaw, K. S., Spracklen, D. V., Mann, G. W. & Woodhouse, M. T. Influence of oceanic dimethyl sulfide emissions on cloud condensation nuclei concentrations and seasonality over the remote Southern Hemisphere oceans: a global model study. *J. Geophys. Res.* **113**, D15204 (2008).
62. Woodhouse, M. T., Mann, G. W., Carslaw, K. S. & Boucher, O. New directions: the impact of oceanic iron fertilization on cloud condensation nuclei. *Atmos. Environ.* **42**, 5728–5730 (2008).
63. Roelofs, G. J. A GCM study of organic matter in marine aerosol and its potential contribution to cloud drop activation. *Atmos. Chem. Phys.* **8**, 709–719 (2008).
64. de Leeuw, G. *et al.* Production flux of sea spray aerosol. *Rev. Geophys.* **49**, 2010RG000349 (2011).
65. O'Dowd, C. D. *et al.* A combined organic-inorganic sea-spray source function. *Geophys. Res. Lett.* **35**, L01801 (2008).
66. Woodhouse, M. T. *et al.* Low sensitivity of cloud condensation nuclei to changes in the sea-air flux of dimethyl-sulphide. *Atmos. Chem. Phys.* **10**, 7545–7559 (2010).
This study modelled the sensitivity of CCN to changes in the sea-to-air flux of DMS and found it to be low, such that the role of DMS in climate regulation is very weak.
67. Carslaw, K. S. *et al.* A review of natural aerosol interactions and feedbacks within the Earth system. *Atmos. Chem. Phys.* **10**, 1701–1737 (2010).
68. Wood, R. Cancellation of aerosol indirect effects in marine stratocumulus through cloud thinning. *J. Atmos. Sci.* **64**, 2657–2669 (2007).
69. Stevens, B. & Feingold, G. Untangling aerosol effects on clouds and precipitation in a buffered system. *Nature* **461**, 607–613 (2009).
70. Small, J. D., Chuang, P. Y., Feingold, G. & Jiang, H. Can aerosol decrease cloud lifetime? *Geophys. Res. Lett.* **36**, L16806 (2009).
71. Zuidema, P., Xue, H. & Feingold, G. Shortwave radiative impacts from aerosol effects on marine shallow cumuli. *J. Atmos. Sci.* **65**, 1979–1990 (2008).
72. Toole, D. A. & Siegel, D. A. Light-driven cycling of dimethylsulfide (DMS) in the Sargasso Sea: closing the loop. *Geophys. Res. Lett.* **31**, L09308 (2004).
73. Vallina, S. M., Simo, R. & Manizza, M. Weak response of oceanic dimethylsulfide to upper mixing shoaling induced by global warming. *Proc. Natl Acad. Sci. USA* **104**, 16004–16009 (2007).
This study modelled the sensitivity of DMS seawater concentrations to a 50% increase in CO₂ and found it to be too low to be a significant offset for global warming.
74. Gunson, J. R. *et al.* Climate sensitivity to ocean dimethyl sulphide emissions. *Geophys. Res. Lett.* **33**, L07701 (2006).
75. Latham, J. & Smith, M. H. Effect on global warming of wind-dependent aerosol generation at the ocean surface. *Nature* **347**, 372–373 (1990).
76. Korhonen, H. *et al.* Aerosol climate feedback due to decadal increases in Southern Hemisphere wind speeds. *Geophys. Res. Lett.* **37**, L02805, doi:10.1029/2009GL041320 (2010).
77. Yang, X.-Y., Huang, R. X. & Wang, D. X. Decadal changes of wind stress over the Southern Ocean associated with Antarctic ozone depletion. *J. Clim.* **20**, 3395–3410 (2007).

Acknowledgements We thank our PhD adviser R. J. Charlson for guidance early in our scientific careers. This review should be seen as 'coming both to praise and bury Caesar' in that the good that the CLAW hypothesis has done will far outlive its use. We also thank W. E. Asher for comments on this manuscript. This is PMEL contribution number 3697.

Author Contributions The ideas presented here were developed jointly by P.K.Q. and T.S.B. and both authors participated actively in the writing of the manuscript and the drafting of the figures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to P.K.Q. (patricia.k.quinn@noaa.gov).

Self-formation of functional adeno-hypophysis in three-dimensional culture

Hidetaka Suga^{1,2,3}, Taisuke Kadoshima¹, Maki Minaguchi¹, Masatoshi Ohgushi², Mika Soen¹, Tokushige Nakano¹, Nozomu Takata¹, Takafumi Wataya¹, Keiko Muguruma¹, Hiroyuki Miyoshi⁴, Shigenobu Yonemura⁵, Yutaka Oiso³ & Yoshiki Sasai¹

The adenohypophysis (anterior pituitary) is a major centre for systemic hormones. At present, no efficient stem-cell culture for its generation is available, partly because of insufficient knowledge about how the pituitary primordium (Rathke's pouch) is induced in the embryonic head ectoderm. Here we report efficient self-formation of three-dimensional adenohypophysis tissues in an aggregate culture of mouse embryonic stem (ES) cells. ES cells were stimulated to differentiate into non-neural head ectoderm and hypothalamic neuroectoderm in adjacent layers within the aggregate, and treated with hedgehog signalling. Self-organization of Rathke's-pouch-like three-dimensional structures occurred at the interface of these two epithelia, as seen *in vivo*, and various endocrine cells including corticotrophs and somatotrophs were subsequently produced. The corticotrophs efficiently secreted adrenocorticotrophic hormone in response to corticotrophin releasing hormone and, when grafted *in vivo*, these cells rescued the systemic glucocorticoid level in hypopituitary mice. Thus, functional anterior pituitary tissue self-forms in ES cell culture, recapitulating local tissue interactions.

The adenohypophysis, corresponding to the anterior and intermediate lobes of the pituitary gland, contains several types of endocrine cells, which secrete systemic hormones such as adrenocorticotrophic hormone (ACTH), GH, LH/FSH, TSH and prolactin^{1–3} in response to hypothalamus-derived releasing hormones. In contrast, the neurohypophysis (posterior pituitary) consists of the axons and secretory termini of hypothalamic vasopressin and oxytocin neurons. During early mammalian development, the adenohypophysis anlage originates as a placode in the non-neural head ectoderm rostral and adjacent to the anterior neural plate^{4,5} (Supplementary Fig. 1a) (hereafter, this part of non-neural ectoderm, expressing region-specific markers *Pitx1* and *Pitx2*, is simply referred to as rostral head ectoderm; this includes the oral ectoderm as a large part). The thickened placodal epithelium invaginates and subsequently detaches from the oral ectoderm, becoming a hollowed epithelial vesicle, Rathke's pouch (Fig. 1a, b and Supplementary Fig. 1b, c). The development of Rathke's pouch depends on tissue interactions between the rostral head ectoderm and the rostral hypothalamus^{3–8}. We sought to recapitulate this interactive microenvironment of the pituitary-forming morphogenetic field using three-dimensional ES cell culture.

Juxtaposing ectodermal layers formed in culture

Mouse ES cells can be induced to differentiate into rostral hypothalamic tissues when cultured as floating aggregates (typically 3,000 cells per aggregate) in a chemically defined medium lacking extrinsic growth factors (including no insulin; hereafter referred to as SFEBQ culture⁹). However, quantitative polymerase chain reaction (qPCR) analysis found only low expression levels of the rostral head ectoderm marker *Pitx2* in SFEBQ-cultured ES cells (Fig. 1c). Because the embryonic anlage for rostral head ectoderm and hypothalamic neuroectoderm are adjacent (Supplementary Fig. 1a–c), we wondered if a slight shift in positional information could promote the simultaneous generation of both tissues within the same aggregate in SFEBQ culture. We experimentally tested several culture conditions known to affect early

ectodermal patterning^{9–12}, and found two that substantially increased *Pitx2* expression levels (Fig. 1d and data not shown): a large cell-aggregation (LCA) culture (red; starting differentiation culture at 10,000 cells per aggregate, instead of 3,000; Supplementary Fig. 2a) and bone morphogenetic protein 4 (BMP4) treatment (blue; 0.5 μ M, days 0–7). BMP4 plus LCA also increased *Pitx2* expression, but no clear additive effects were seen over LCA alone (Supplementary Fig. 2b).

In the LCA culture, both *Pitx1/2*⁺ rostral head ectoderm tissue and *Rx*⁺ neural tissue formed continuous epithelia within the ES cell aggregates (Fig. 1e–g; Supplementary Fig. 2c and data not shown). The *Rx*⁺ tissues were *Chx10*[–] *nestin*⁺ (*Chx10* also known as *Vsx2*; Supplementary Fig. 2d), characteristic of rostral hypothalamic epithelia (retinal epithelia are *Rx*⁺ but *Chx10*⁺ *nestin*[–])^{9,13}. Importantly, a layer of *Pitx1/2*⁺ epithelium reproducibly formed on the surface of the ES cell aggregates, adjacent and exterior to the *Rx*⁺ neuroepithelia, which formed inner layers (Fig. 1f, g). In contrast, treatment with exogenous BMP4, which is generally known to promote differentiation of non-neural ectoderm at the cost of neural differentiation^{10,12}, inhibited the formation of hypothalamic tissues even at 0.5 nM, regardless of the aggregation size (Supplementary Fig. 2e). Conversely, treatment with the BMP antagonist dorsomorphin inhibited the generation of *Pitx2*⁺ ectoderm, which was promoted in the LCA culture (Supplementary Fig. 2f). Given that the expression of *Bmp2* and *Bmp4* (refs 8, 14) was increased under LCA conditions (Supplementary Fig. 2g–i), the moderate elevation of endogenous BMP signals in the LCA culture, which was insufficient to inhibit hypothalamic differentiation, may have contributed to the spontaneous formation of rostral head ectoderm.

Self-formation of Rathke's pouches *in vitro*

In the embryo, Rathke's pouch develops on the midline of the rostral head ectoderm, which receives strong sonic hedgehog (Shh) signals^{3–7,15} (Supplementary Fig. 1d). Accordingly, treatment with the

¹Neurogenesis and Organogenesis Group, RIKEN Center for Developmental Biology, Kobe 650-0047, Japan. ²Division of Human Stem Cell Technology, RIKEN Center for Developmental Biology, Kobe 650-0047, Japan. ³Department of Endocrinology and Diabetes, Graduate School of Medicine, Nagoya University, Nagoya, 466-8550, Japan. ⁴Subteam for Manipulation of Cell Fate, BioResource Center, RIKEN, Tsukuba 305-0074, Japan. ⁵Electron Microscopy Laboratory, RIKEN Center for Developmental Biology, Kobe 650-0047, Japan.

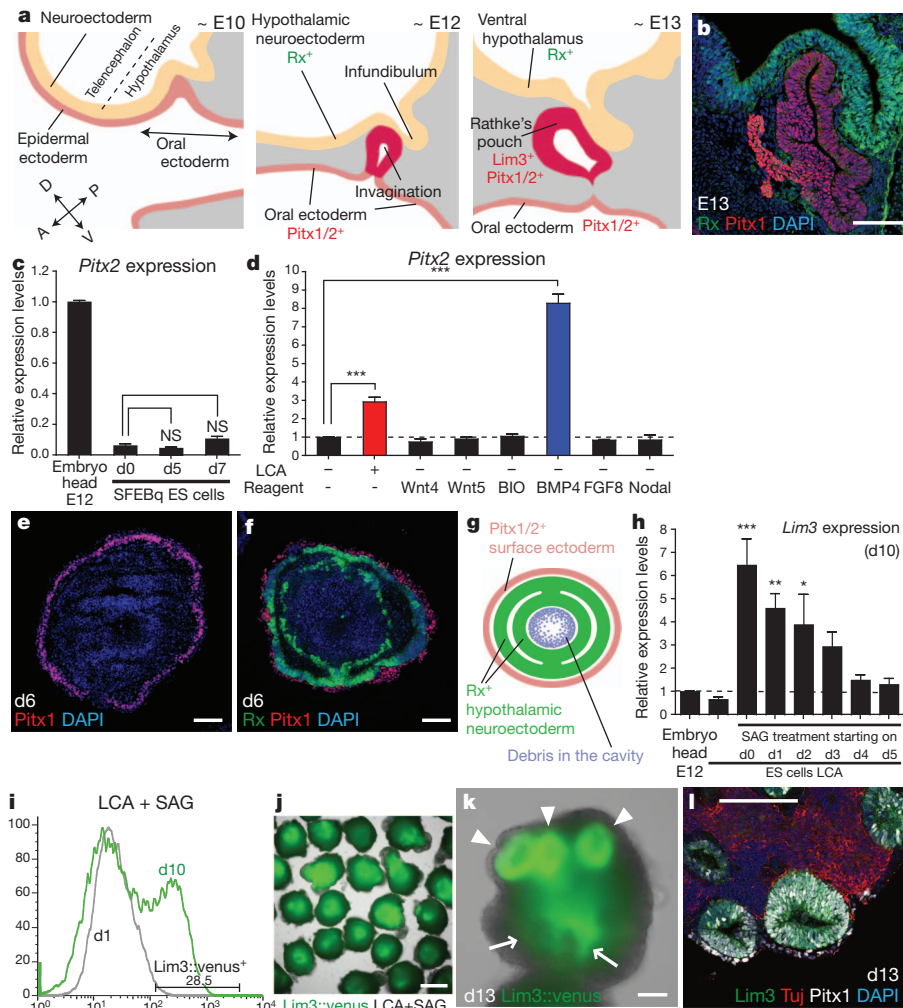


Figure 1 | Adjacent formation of head ectoderm and hypothalamic epithelia in ES cell culture. **a**, Schematic of pituitary development, sagittal view. A, anterior; D, dorsal; P, posterior; V, ventral. **b**, *In vivo* expression of Pitx1 and Rx on embryonic day 13 (E13). **c**, Little expression of *Pitx2* in SFEBq/CDM-cultured ES cells (qPCR; $n = 3$ experiments). **d**, day. **d**, *Pitx2* expression in aggregates cultured under the LCA condition (red) or treated with 0.5 nM BMP4 (blue) ($n = 3$). **e**, **f**, Immunostaining of day-6 LCA aggregates for Pitx1 (red), and Rx (green). **g**, Schematic of the result. **h**, Induction of *Lim3*

expression by SAG (qPCR; $n = 9$). **i**, FACS analysis of the *Lim3::venus*⁺ population in day-10 LCA aggregates with SAG treatment (green). Grey, day-1 aggregate. **j**, *Lim3::venus* expression in LCA aggregates with SAG treatment. **k**, Clusters of *Lim3::venus*⁺ vesicles (arrowheads) in day-13 SAG-treated LCA aggregates. Arrows, vesicles forming on the other side. **l**, *Lim3*⁺ pouch is *Pitx1*⁺ but *Tuj1*⁻. Scale bars, 100 μ m (**b**, **e**, **f**, **k**, **l**). The values shown on graphs represent the mean \pm s.e.m. NS, not significant. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

hedgehog (Hh) agonist N-methyl-N'-(3-pyridinylbenzyl)-N'-(3-chlorobenzo[b]thiophene-2-carbonyl)-1,4-diaminocyclohexane (smoothened agonist; SAG, hereafter)¹⁶ (days 0–10) efficiently increased expression of the Rathke's pouch marker^{17,18} *Lim3* (also called *Lhx3*) in LCA-SFEBq culture (Fig. 1h). In a culture using venus-reporter ES cells for *Lim3* expression (*Lim3*^{venus/+}; Supplementary Fig. 3a, b), SAG reproducibly induced *Lim3::venus*⁺ cells in >90% of LCA-SFEBq aggregates (25–30% of total cells on day 10; Fig. 1i, j and Supplementary Fig. 3c). Additional BMP4 treatment (SAG+BMP4 under the LCA condition) had no additive effect (rather, an inhibitory effect; Supplementary Fig. 3d) on *Lim3::venus* expression.

In SAG-treated LCA-SFEBq aggregates, the *Lim3::venus*⁺ cells formed multiple oval epithelial clusters (Fig. 1k and Supplementary Movie 1; unlike *Lim3*⁺ neurons in the brain, they were *Tuj1*⁻; Fig. 1l). No concomitant induction of the mesodermal markers brachyury and Flk1 (also known as *Kdr*) or the neural crest marker *Foxd3* was observed in the aggregates (Supplementary Fig. 4a–d).

These *Lim3::venus*⁺ tissues first appeared as thickened placode epithelium on the surface (~day 8), invaginated (days 9–10) and formed hollowed epithelial vesicles (days 10–11), as in embryonic

Rathke's pouch development (Fig. 2a–c and Supplementary Fig. 4e). The *Lim3::venus*⁺ epithelium co-expressed *Pitx1*, *Pitx2* and E-cadherin, and was negative for neuroepithelial markers such as nestin, Rx and *Nkx2.1* or the surface ectoderm marker cytokeratin (Fig. 2d–g and Supplementary Fig. 4f–h), reminiscent of the marker profile of the pituitary primordium^{1–3,19} (Supplementary Fig. 4i).

The *Lim3::venus*⁺ epithelial vesicles were reproducibly located between the hypothalamic (*Rx*⁺) and the rostral head ectodermal (*Pitx1*⁺/*Lim3*⁻) tissues, and the atypical protein kinase C (aPKC)⁺ apical surface of the *Lim3::venus*⁺ epithelial vesicle was inside along the apical–basal axis (Supplementary Fig. 4j, k), consistent with the topographical development of Rathke's pouch *in vivo*. The pouch epithelia consisted of tall columnar epithelial cells with apical cilia and junctions (Fig. 2h and Supplementary Fig. 5), and the late-onset Rathke's pouch marker^{1,6,20} *islet1/2* (Fig. 2i) was expressed in a differentiating cell population located basally on day 13 (see Fig. 2h, bracket). On days 13–14, the long axis of the *Lim3*⁺ pouch typically reached a diameter of 150–200 μ m, comparable to that of early Rathke's pouch *in vivo*.

Thus, in the presence of high Hh signals, the ES-cell-derived rostral head ectoderm adjacent to the hypothalamic neuroepithelium initiates

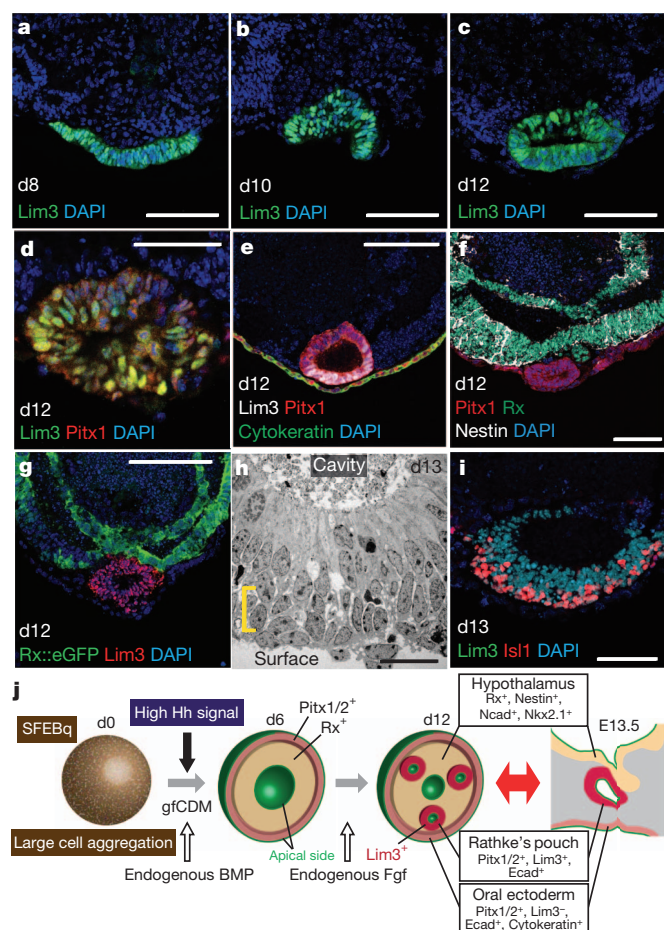


Figure 2 | Spontaneous generation of Rathke's pouch-like vesicles in ES cell culture. **a–c**, Morphogenesis of Lim3^+ epithelia. **d–g**, Immunostaining of day-12 pouch vesicles and surrounding tissues for Pitx1 (red, **d–f**), Lim3 (green, **d**; white, **e**; red, **g**), pancytokeratin (green, **e**), nestin (white, **f**) and Rx (green, **f**, **g**) in ES cell culture. **h**, Electron microscopy of the day-13 pouch. Delaminating cells on the basal side (bracket). **i**, Islet1^+ cells in the basal zone of the day-13 pouch. **j**, Schematic of *in vitro* generation of Rathke's pouches. Scale bars, 100 μm (**a–c**, **e–g**); 50 μm (**d**, **i**); 20 μm (**h**).

Rathke's pouch development in three-dimensional culture (Fig. 2j). The interactions between the two juxtaposed tissues seemed to be critical, as Pitx1/2^+ surface epithelia lacking Rx^+ neural tissues in BMP4-treated aggregates (Fig. 1d and Supplementary Fig. 2e) did not exhibit visible pouch development, regardless of Hh signals (data not shown). In addition, surface ectoderm isolated alone from day-6 aggregates exhibited low Lim3 expression by qPCR when cultured until day 13, unless combined with hypothalamic neuroepithelium (Supplementary Fig. 6a–c).

Fgf signals such as Fgf10 are implicated in Rathke's pouch development^{6–8,21}. Consistent with this idea, the Fgf receptor inhibitors SU5402 and PD173074 reduced the percentages of $\text{Lim3}::\text{venus}^+$ cells in culture (Supplementary Fig. 6d). Fgf10 expression was gradually induced in the Rx^+ tissues of the aggregate during days 7–10, and Fgf8 expression was observed in both Rx^+ and Rx^- populations (Supplementary Fig. 6e, f). phospho-Erk (pErk), which is indicative of Fgf signalling, predominantly accumulated in the Pitx1^+ surface epithelia (Supplementary Fig. 6g). This preferential pErk accumulation may be relevant to the localized expression of *FgfR2b* (Fgf10-specific receptor gene²²) and *FgfR4* (receptor gene for Fgf8) in the Rx^- tissues (Supplementary Fig. 6h, i). These observations indicate that endogenous Fgf signals also have a role in Lim3 induction, at least in part, in this three-dimensional ES cell culture.

Generation of multiple endocrine lineages

During early pituitary development²³, Lim3^+ pituitary progenitors become committed to several hormone-type-specific lineages^{1–6} (Fig. 3a; Supplementary Fig. 7a for details). Among them, the ACTH-producing corticotroph lineage requires the transcription factor *Tbx19* (ref. 24), the expression of which is inhibited by Notch signalling^{25,26} (Fig. 3a). Consistent with this idea, whereas *Tbx19* expression in SAG-treated LCA-SFEbq aggregates was low, it was strongly increased by treatment with the Notch inhibitor (2S)-N-[N-(3,5-Difluorophenacetyl)-L-alanyl]-2-phenylglycine tert-butyl ester (DAPT) (Fig. 3b and Supplementary Fig. 7b). Substantial numbers of large-sized ACTH⁺ cells reproducibly accumulated in the Tbx19^+ domains of DAPT-treated pouch tissue (day 21; Fig. 3d and Supplementary Fig. 7c); ACTH⁺ cells comprised ~35% of N-cadherin[−] non-neural cells (Fig. 3c), which themselves represented ~11% of total cells on day 21. The ACTH⁺ cells were negative for Ki67 (not shown) and E-cadherin (Supplementary Fig. 7d), unlike the early progenitors in the pouch (for example, Supplementary Fig. 4h). The ES-cell-derived ACTH⁺ cells were negative for the neuronal markers neurofilament and NSE (Fig. 3e and Supplementary Fig. 7e) and the melanotroph-lineage marker²⁴ PC2 (Fig. 3f and Supplementary Fig. 7f), consistent with the corticotroph nature of marker expression.

Lim3 is essential for early pituitary specification^{17,18,27}. Knockdown of *Lim3* (short hairpin RNA (shRNA)-based; Supplementary Fig. 7g) specifically reduced expression of *Tbx19* and *Pomc* (mouse homologue of human *ACTH*, hereafter referred to as the latter) in culture (Fig. 3g; Supplementary Fig. 7h for controls). Similarly, when *Tbx19*^{−/−} ES cells (Supplementary Fig. 7i) were used, *ACTH* expression was significantly reduced (Fig. 3h). Thus, LCA-SFEbq culture with SAG treatment faithfully recapitulates *in vivo* corticotroph development, which is inhibited by Notch and driven by *Lim3* and *Tbx19* (Fig. 3i).

In the developing pituitary, Tbx19^- progenitors give rise to Pit1^+ intermediate precursors, which subsequently differentiate into GH-, prolactin- and TSH-producing cell lineages^{3–6,28} (Fig. 3a). Unlike *Tbx19* expression, *Pit1* expression in the SAG-treated aggregates was not increased by DAPT treatment (Fig. 3j). In contrast, consistent with previous reports that canonical Wnt signalling promotes *Pit1* expression²⁸, its expression was specifically augmented by treatment with the Wnt agonist 6-bromo-3-[(3E)-1,3-dihydro-3-(hydroxyimino)-2H-indol-2-ylidene]-1,3-dihydro-(3Z)-2H-indol-2-one (BIO) (GSK3 β inhibitor) on days 16–18 (Fig. 3j; see also Fig. 3b, c). Further differentiation of these intermediate precursors into GH⁺ and prolactin⁺ cells was observed (Fig. 3k–n) when cultured in media containing glucocorticoid and oestradiol, respectively (Supplementary Fig. 8a, b), which enhance the differentiation of these lineages^{29,30}.

Head mesenchyme has been suggested to have a promoting role in pituitary development³¹. With this in mind, we treated LCA-SAG aggregates with conditioned medium of PA6 stromal cells (derived from skull bone-marrow mesenchyme) (Supplementary Fig. 8c). This treatment enabled the generation of LH⁺ and FSH⁺ cells (also some TSH⁺ cells; Fig. 3o–r), which was otherwise very rare (data not shown).

These findings demonstrate the competence of ES-cell-derived adenohypophysis progenitors to generate multiple endocrine lineages in culture.

In vitro functionality as endocrine tissue

To investigate the *in vitro* hormone-secretion ability of adenohypophysis-like tissue generated *in vitro*, we focused on corticotrophs, which were most efficiently generated in this three-dimensional ES cell culture. Substantial ACTH release from SAG+DAPT-treated LCA aggregates was induced after a 10-min stimulation with corticotrophin releasing hormone (CRH) on day 21 (Fig. 4a, b and Supplementary Fig. 9a), but not from those without SAG treatment (Fig. 4c). The aggregates treated with SAG but not with DAPT showed ACTH secretion upon CRH

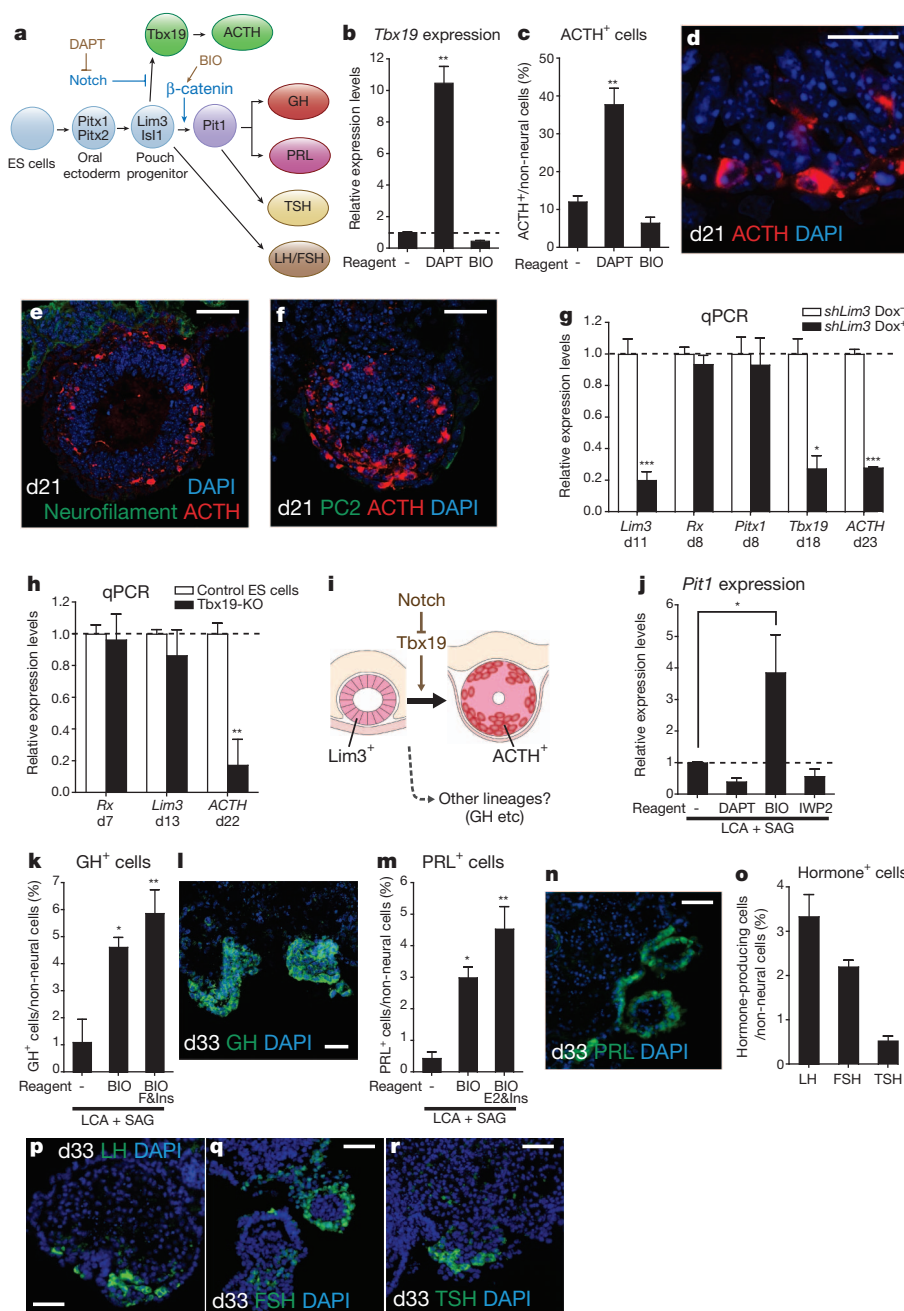


Figure 3 | Differentiation of ES-cell-derived progenitors into hormone-producing cell lineages. **a**, Schematic of lineage-specific differentiation. **b**, Increased *Tbx19* expression on day 20 by DAPT (days 18–19). BIO treatment (days 16–18). qPCR analysis ($n = 3$ experiments). **c**, Percentages of *ACTH*⁺ cells in non-neural (N-cad⁻) cells on day 22 ($n = 6$). **d–f**, Immunostaining of day-21 SAG+DAPT-treated aggregates for *ACTH* (red), neurofilament (green, **d**) and PC2 (green, **e**). **g**, Attenuated *Tbx19* and *ACTH* expression by doxycycline-induced *Lim3* knockdown (Dox; black columns) ($n = 6$). **h**, Expression of *Rx* (day 7), *Lim3* (day 13) and *ACTH* (day 22) by qPCR analysis of LCA+SAG+DAPT-treated aggregates ($n = 3$). Black columns, *Tbx19* knockout (*Tbx19*-KO) ES cells. **i**, Schematic of corticotroph

generation *in vitro*. **j**, *Pit1* expression in day-26 culture of LCA+SAG aggregates ($n = 3$). **k–o**, Non-corticotroph differentiation. **k**, **m**, Percentages of GH⁺ (**k**) and prolactin⁺ (PRL; **m**) cells in non-neural cells of day-33 LCA+SAG aggregates. F&Ins, hydrocortisone + insulin treatment; E2&Ins, oestradiol + insulin treatment. **l**, **n**, Immunostaining of day-33 aggregates for GH (**l**) and prolactin (**n**) (green). **o**, Percentages of non-neural cells producing LH, FSH and TSH in day-33 aggregates cultured with PA6 conditioned medium during days 10–33. **p–r**, LH, FSH and TSH immunostaining on day 33. Scale bars, 20 μ m (**d**); 50 μ m (**e**, **f**); 100 μ m (**l**, **n**, **p–r**). The values shown on graphs represent the mean \pm s.e.m. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

stimulation, but significantly less than SAG+DAPT-treated aggregates did (Fig. 4c).

ACTH secretion from the pituitary is negatively regulated by the downstream hormone glucocorticoid³² (Fig. 4d, inset). Consistent with this *in vivo* regulation, the *in vitro* release of ACTH by CRH was greatly suppressed by pre-treating with glucocorticoid (Fig. 4d), but not with oestrogen (Supplementary Fig. 9b). These findings demonstrate that the ES-cell-derived pituitary endocrine cells actively

secrete ACTH and respond normally to both positive and negative regulators that work for endocrine homeostasis *in vivo*.

Functional rescue in hypopituitary mice

Lastly, we evaluated ACTH secretion *in vivo* using hypophysectomized mice³³ (Fig. 5a; see the reduction of blood ACTH levels in Supplementary Fig. 9c, d). In mouse, the pituitary is located at a tiny space in the skull base, called the sella turcica^{3–6}. Because its local circulatory

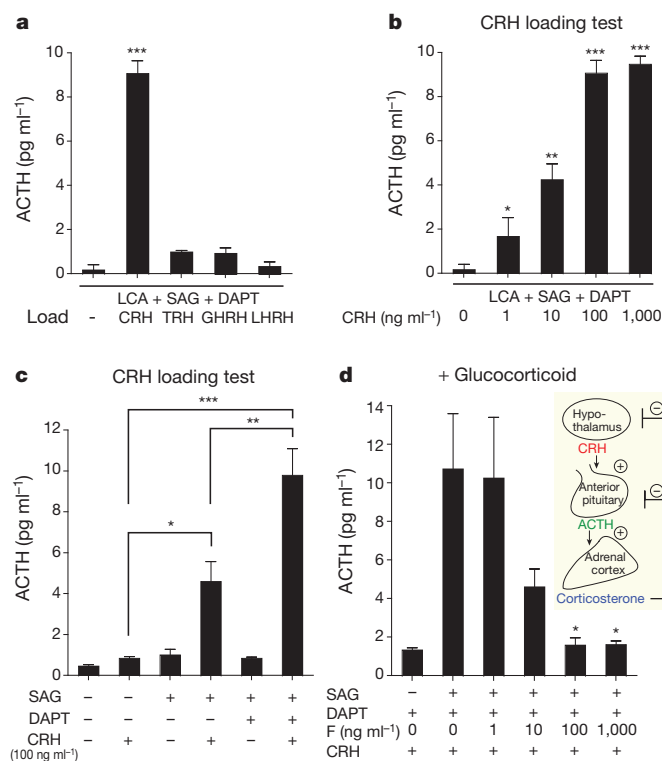


Figure 4 | CRH-induced ACTH release from ES-cell-derived pituitary tissues. **a**, CRH efficiently induced ACTH secretion ($n = 4$ experiments). **b**, ACTH secretion at different CRH doses ($n = 4$). **c**, ACTH secretion from aggregates cultured in different conditions ($n = 3$). **d**, Pre-treatment with hydrocortisone ($n = 3$) suppressed the CRH-stimulated ACTH secretion from aggregates. The values shown on graphs represent the mean \pm s.e.m. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$.

network including hypophyseal portal veins is easily destroyed by hypophysectomy, we performed ectopic transplantation of aggregates under the kidney capsules in this study (Fig. 5a; see grafted ACTH⁺ cells stained in red).

A week after transplantation, CRH loading induced a substantial elevation of blood ACTH levels in the mice grafted with SAG+DAPT-treated aggregates (Fig. 5b) but not in control hypophysectomized mice that had received aggregates without SAG/DAPT treatment (Fig. 5c). Accordingly, the blood glucocorticoid (corticosterone) levels were increased upon CRH loading in the mice grafted with SAG+DAPT-treated aggregates but not in controls (Fig. 5d), indicating that ACTH from the graft sufficiently induced the downstream hormone.

Even without CRH loading, the basal levels of ACTH were higher in the mice that received SAG+DAPT-treated aggregates than in controls (Fig. 5e). Importantly, the corticosterone levels were also increased (Fig. 5f), suggesting that this partial recovery of blood ACTH has a moderate but biologically significant effect (note that the median effective dose (ED₅₀) of the ACTH receptor MC2R for glucocorticoid production is around 9 pg ml⁻¹; ref. 34). In accordance with this idea, the hypophysectomized mice receiving SAG+DAPT-treated aggregates showed higher spontaneous locomotor activity in their home cages than controls (infrared-monitored 24-h tracking in Fig. 5g; also examined by the running wheel test; Fig. 5h and Supplementary Movie 2; Supplementary Fig. 9e, f for controls), and survived significantly longer ($P < 0.01$, Kaplan–Meyer method; Fig. 5i).

Conclusion and perspective

Although the adenohypophysis is relatively small in size, it is an independent endocrine organ with indispensable systemic functions. Using a synthetic approach at the multi-cellular level, we demonstrated that a functional organ bud, the development of which requires intricate

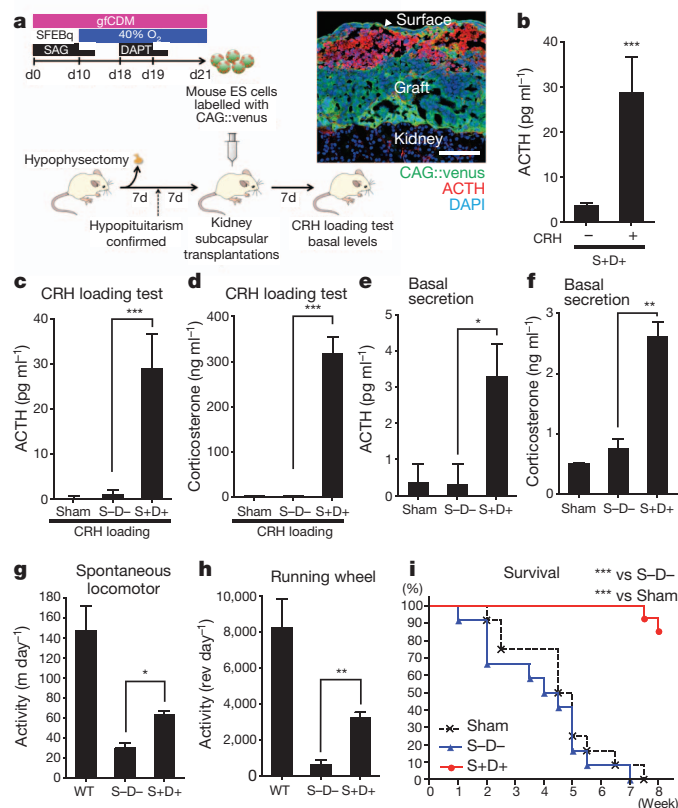


Figure 5 | Increased systemic ACTH and glucocorticoid levels by grafting ES-cell-derived pituitary tissues. **a**, Schematic of transplantation procedures. Inset, ACTH⁺ cells (red) in grafted tissue (labelled with GFP; post-transplantation day 7) under the renal capsule. **b**, Blood ACTH levels in grafted mice without SAG or DAPT treatment; S+D+, aggregates with SAG+DAPT treatment. **c**, **d**, Blood ACTH (**c**) and corticosterone (**d**) levels upon CRH loading (sampled two hours after loading; 7 days after transplantation; $n = 3$). Sham, subcapsular saline injection. **e**, **f**, Basal blood levels of ACTH (**e**) and corticosterone (**f**) 7 days after transplantation ($n = 3$). No CRH loading. Blood sampling at 20:00 in the evening. **g**, Spontaneous locomotor activity (moving distance) in the home cage measured for 24 h by the infrared monitoring system. WT, wild type. Control ($n = 8$) versus S+D+ ($n = 7$), $P < 0.05$, Bonferroni's test. Wild type (no hypophysectomy), $n = 4$. **h**, Locomotor measured by running wheels. Control ($n = 9$) versus S+D+ ($n = 7$), $P < 0.01$. Wild type, $n = 4$. **i**, Improved survival of transplanted hypopituitary mice. S-D-, $n = 12$; S+D+, $n = 14$. Sham, $n = 12$. The values shown on graphs represent the mean \pm s.e.m. *, $P < 0.05$; **, $P < 0.01$; ***, $P < 0.001$. Scale bar, 100 μ m (**a**).

tissue interactions, can be reproducibly generated by self-formation when the spatial arrangement of a few interacting tissues is reconstituted in three-dimensional culture. In terms of mechanistic complexity, this idea represents a further conceptual advance beyond our recent report of optic-cup self-organization, which sufficiently occurs within retinal epithelium alone¹³.

Hypopituitarism is a major category of endocrinological disorders, including empty sella syndrome, Sheehan syndrome, hypophyseal apoplexy and post-operative pituitary damage^{1,2}. In addition, insufficiencies of specific hormones (for example, GH insufficiency in infants and adults) also cause various systemic problems. The successful generation of pituitary endocrine cells shown here opens new avenues for the application of pluripotent stem cells to treat non-diabetic endocrine diseases, which have so far received relatively little attention in regenerative medicine. A future challenge will be orthotopic transplantation of adenohypophysis derived from human ES/induced pluripotent stem cells into the sella of a larger mammal, together with local reconstitution of vascular networks.

METHODS SUMMARY

The SFEBq culture was performed as described previously except that cells were aggregated at 10,000 cells per aggregate in 96-well plates and cultured in growth-factor-free chemically defined medium (gfCDM) medium supplemented with 400 nM SAG on days 0–10 unless otherwise stated. For endocrine cell differentiation, culture was continued under 40% O₂. Surgical hypophysectomy was performed by the transaural approach³³.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 26 July; accepted 14 October 2011.

Published online 9 November 2011.

- Kelberman, D., Rizzoti, K., Lovell-Badge, R., Robinson, I. C. & Dattani, M. T. Genetic regulation of pituitary gland development in human and mouse. *Endocr. Rev.* **30**, 790–829 (2009).
- Romero, C. J., Nesi-França, S. & Radovick, S. The molecular basis of hypopituitarism. *Trends Endocrinol. Metab.* **20**, 506–516 (2009).
- Rizzoti, K. & Lovell-Badge, R. Early development of the pituitary gland: induction and shaping of Rathke's pouch. *Rev. Endocr. Metab. Disord.* **6**, 161–172 (2005).
- Takuma, N. *et al.* Formation of Rathke's pouch requires dual induction from the diencephalon. *Development* **125**, 4835–4840 (1998).
- Suh, H., Gage, P. J., Drouin, J. & Camper, S. A. Pitx2 is required at multiple stages of pituitary organogenesis: pituitary primordium formation and cell specification. *Development* **129**, 329–337 (2002).
- Zhu, X., Gleiberman, A. S. & Rosenfeld, M. G. Molecular physiology of pituitary development: signaling and transcriptional networks. *Physiol. Rev.* **87**, 933–963 (2007).
- Rizzoti, K. & Lovell-Badge, R. Early development of the pituitary gland: induction and shaping of Rathke's pouch. *Rev. Endocr. Metab. Disord.* **6**, 161–172 (2005).
- Ericson, J., Norlin, S., Jessell, T. M. & Edlund, T. Integrated FGF and BMP signaling controls the progression of progenitor cell differentiation and the emergence of pattern in the embryonic anterior pituitary. *Development* **125**, 1005–1015 (1998).
- Wahaya, T. *et al.* Minimization of exogenous signals in ES cell culture induces rostral hypothalamic differentiation. *Proc. Natl Acad. Sci. USA* **105**, 11796–11801 (2008).
- Watanabe, K. *et al.* Directed differentiation of telencephalic precursors from embryonic stem cells. *Nature Neurosci.* **8**, 288–296 (2005).
- Eiraku, M. *et al.* Self-organized formation of polarized cortical tissues from ESCs and its active manipulation by extrinsic signals. *Cell Stem Cell* **3**, 519–532 (2008).
- Basch, M. L. & Bronner-Fraser, M. Neural crest inducing signals. *Adv. Exp. Med. Biol.* **589**, 24–31 (2006).
- Eiraku, E. *et al.* Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature* **472**, 51–56 (2011).
- Jones, C. M., Lyons, K. M. & Hogan, B. L. Involvement of bone morphogenetic protein-4 (BMP-4) and Vgr-1 in morphogenesis and neurogenesis in the mouse. *Development* **111**, 531–542 (1991).
- Danjo, T. *et al.* Subregional specification of ES cell-derived ventral telencephalic tissues by timed and combinatory treatment with extrinsic signals. *J. Neurosci.* **31**, 1919–1933 (2011).
- Treier, M. *et al.* Hedgehog signaling is required for pituitary gland development. *Development* **128**, 377–386 (2001).
- Ellsworth, B. S., Butts, D. L. & Camper, S. A. Mechanisms underlying pituitary hypoplasia and failed cell specification in Lhx3-deficient mice. *Dev. Biol.* **313**, 118–129 (2008).
- Sheng, H. Z. *et al.* Specification of pituitary cell lineages by the LIM homeobox gene Lhx3. *Science* **272**, 1004–1007 (1996).
- Kikuchi, M. *et al.* Changes in E- and N-cadherin expression in developing rat adenohypophysis. *Anat. Rec.* **290**, 486–490 (2007).
- Thor, S., Ericson, J., Brännström, T. & Edlund, T. The homeodomain LIM protein Isl-1 is expressed in subsets of neurons and endocrine cells in the adult rat. *Neuron* **7**, 881–889 (1991).
- Bharti, K., Gasper, M., Bertuzzi, S. & Arnheiter, H. Lack of the ventral anterior homeodomain transcription factor VAX1 leads to induction of a second pituitary. *Development* **138**, 873–878 (2011).
- Mason, I. Initiation to end point: the multiple roles of fibroblast growth factors in neural development. *Nature Rev. Neurosci.* **8**, 583–596 (2007).
- Davis, S. W., Mortensen, A. H. & Camper, S. A. Birthdating studies reshape models for pituitary gland cell specification. *Dev. Biol.* **352**, 215–227 (2011).
- Lamolet, B. *et al.* A pituitary cell-restricted T box factor, Tpit, activates POMC transcription in cooperation with Pitx homeoproteins. *Cell* **104**, 849–859 (2001).
- Zhu, X. *et al.* Sustained Notch signaling in progenitors is required for sequential emergence of distinct cell lineages during organogenesis. *Genes Dev.* **20**, 2739–2753 (2006).
- Kita, A. *et al.* Hes1 and Hes5 control the progenitor pool, intermediate lobe specification, and posterior lobe formation in the pituitary development. *Mol. Endocrinol.* **21**, 1458–1466 (2007).
- Zhao, Y. *et al.* Reduced expression of the LIM-homeobox gene Lhx3 impairs growth and differentiation of Rathke's pouch and increases cell apoptosis during mouse pituitary development. *Mech. Dev.* **123**, 605–613 (2006).
- Davis, S. W. *et al.* Molecular mechanisms of pituitary organogenesis: in search of novel regulatory genes. *Mol. Cell. Endocrinol.* **323**, 4–19 (2010).
- Hemming, F. J., Bégeot, M., Dubois, M. P. & Dubois, P. M. Fetal rat somatotropes *in vitro*: effects of insulin, cortisol, and growth hormone-releasing factor on their differentiation: a light and electron microscopic study. *Endocrinology* **114**, 2107–2113 (1984).
- Ogasawara, K. *et al.* Hormonal regulation of prolactin cell development in the fetal pituitary gland of the mouse. *Endocrinology* **150**, 1061–1068 (2009).
- Gleiberman, A. S., Fedtsova, N. G. & Rosenfeld, M. G. Tissue interactions in the induction of anterior pituitary: role of the ventral diencephalon, mesenchyme, and notochord. *Dev. Biol.* **213**, 340–353 (1999).
- Ingle, D. J. The effects of administering large amounts of cortin on the adrenal cortices of normal and hypophysectomized rats. *Am. J. Physiol.* **124**, 369–371 (1938).
- Falconi, G. & Rossi, G. L. Transauricular hypophysectomy in rats and mice. *Endocrinology* **74**, 301–303 (1964).
- Melmed, S. *The Pituitary* 3rd edn, 61 (Academic, 2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We are grateful to H. Enomoto, R. Ladher and M. Eiraku for invaluable comments, to K. Misaki for electron microscopy analysis, and to members of the Y.S. laboratory for discussion. This work was supported by grants-in-aid from Ministry of Education, Culture, Sports, Science and Technology (Y.S., Y.O.), the Knowledge Cluster Initiative at Kobe, and the Leading Project for Realization of Regenerative Medicine (Y.S.).

Author Contributions H.S. and Y.S. designed the project and wrote the manuscript. H.S., T.K., M.O. and M.M. performed the experiments with the technical help and advice of T.N., N.T., M.S., K.M., H.M., S.Y. and T.W., and Y.O. provided critical advice on the research strategy and design.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Y.S. (yoshikisai@cdb.riken.jp).

METHODS

ES cell culture for maintenance and differentiation. Mouse ES cells (EB5, *Rx::EGFP, Sox1::GFP*) were maintained as described, in G-MEM supplemented with 1% fetal calf serum, 10% KSR, 2 mM glutamine, 0.1 mM non-essential amino acids, 1 mM pyruvate, 0.1 mM 2-mercaptoethanol and 2,000 U ml⁻¹ LIF. For SFEBq culture, ES cells were dissociated to single cells in 0.25% trypsin-EDTA, and quickly reaggregated in differentiation medium (10,000 cells per 100 µl per well unless otherwise stated) using 96-well low-cell-adhesion plates (Lipidure Coat, NOF Corp). The differentiation medium was growth-factor-free CDM (gfCDM), which contains Iscove's modified Dulbecco's medium/Hams F12 1:1, 1 × chemically defined lipid concentrate, monothioglycerol (450 µM) and 5 mg ml⁻¹ purified bovine serum albumin (>99% purified by crystallization; Sigma). Growth factors were purchased from R&D Systems and used at the concentration indicated in the text and legends. The concentrations and durations of the reagents applied to differentiation culture were as follows: DAPT (10 µM, days 18–19), SAG (400 nM, days 0–10), BIO for Fig. 1 (250 nM, days 0–7) and Fig. 3 (250 nM, days 16–18), BMP4 (0.5 nM, days 0–7), dorsomorphin (1 µM, days 0–7), Wnt4 and Wnt5 (50 ng ml⁻¹, days 0–7, each), Fgf8 (200 ng ml⁻¹, days 0–7), Nodal (1 µg ml⁻¹, days 0–7) and IWP2 (2 µM, days 16–18). The day when SFEBq culture was initiated was defined as day 0. For LCA culture, aggregation of different numbers of ES cells on day 0 was tested as follows: 3,000, 5,000, 10,000, 15,000 and 20,000 cells per well. Among them, efficient generation of Pitx1⁺ surface ectoderm was reproducibly observed in aggregates with 10,000, 15,000 and 20,000 cells. However, culture of too large aggregates (in particular, 20,000-cell aggregates) tended to be less healthy and Lim3:venus induction in them was less efficient than in those with 10,000 ES cells. Therefore, the condition of 10,000 cells per well was used in this study. In the aggregate with this starting cell number, the innermost portion was either a cavity or mass of cell debris (more frequently the latter) on day 10 (Fig. 1g).

For isolation culture of surface ectoderm (Supplementary Fig. 6a–c), outer epithelium of day-6 aggregates were manually detached by forceps and a tungsten needle and cultured alone or together with inner neuroepithelium in SAG-containing differentiation medium. Fgf10 treatment alone (regardless of SAG treatment) was not sufficient to induce *Lim3* expression in the surface epithelium from aggregates, indicating the necessity of additional unknown inductive signals.

Generation of knock-in ES cell lines. The gene-targeting strategy and vector construction for *Lim3::venus* is as illustrated in Supplementary Fig. 2a. ES cells (EB5) were cotransfected with the targeting vector and an expression vector carrying the zinc-finger nuclease (ZFN) for exon 1 of *Lim3b* (*Lim3* has two mRNA subtypes, each using a different exon 1; ZFN targets 5'-GACCTGCTGCTG-3' and 5'-TTGGCGCGAAGGGCTGAC-3'), and subjected to G418 selection. Colonies knocked-in at one allele were genotyped with PCR, Southern blotting and genomic sequencing. To generate the targeting construct, we used the MultiSite Gateway Technology (Invitrogen). Briefly, the att site-flanked PCR products of the 5' arm (1.0 kbp) and 3' arm (1.0 kbp) were cloned into *pDONR* vector. The *venus* cDNA was fused in-frame into exon 1b of the *Lim3* gene at the initial ATG by using Gateway LR clonase (Invitrogen). A *PGK* promoter-driven neomycin-resistance selection cassette flanked by *loxP* sites was inserted downstream of *venus*. Homologous recombinant ES cells were selected with neomycin. Three targeted clones (#K2, #J2 and #J4) were confirmed by genomic PCR and sequencing (Supplementary Fig. 2a and data not shown) and exhibited indistinguishable abilities to differentiate into Lim3⁺ progenitors; subclone #J4 was mainly used for the experiments shown in this report. A similar strategy was used for targeting *Tbx19*; the vector construction is illustrated in Supplementary Fig. 5g. The ZFN target sequences for exon1 were 5'-GTCTCGCTGCTTAAC-3' and 5'-GAAAGTGAGCTCCAG-3'. The clone in which both *Tbx19* alleles were disrupted was used for the differentiation assay.

shRNA-mediated knockdown study. To suppress *Lim3* expression in differentiating cells, we used the Tet-inducible shRNA expression lentivirus vector system (invented by H.M. at the RIKEN BioResource Center). *pENTR4-H1tetOx1*, in which the *tet* operator sequence (5'-TCCCTATCAGTGATAGAGA-3') was inserted between the TATA box in *H1* promoter and the transcription start site, was designed to express shRNA under the control of the Tet-inducible *H1* promoter (*H1tetO*). The shRNA sequences targeting mouse *Lim3* mRNA, 5'-GCACATCTTGACCGTTTCAT-3' (A), 5'-CGGCAGTTCGAAGTCCGACAA-3' (B) and 5'-CGCCTACAACACTTCGCCCAA-3' (C), were inserted into *pENTR4-H1tetOx1*. As a negative control, the shRNA sequence targeting *E. coli lacZ* mRNA, 5'-GACTACACAAATCAGCGATT-3' (D), was also inserted. The Tet-inducible shRNA expression cassette in *pENTR4-H1tetOx1* was transferred to the *CS-RfA-ETR* lentivirus vector, which contains an expression cassette for tetracycline repressor (TetR) and monomeric red fluorescent protein (mRFP1) linked with a 2A-self-cleaving peptide sequence under the

control of the human *EF1alpha* promoter, by using Gateway LR Clonase (Invitrogen).

To produce recombinant lentiviruses, the transfer vector was co-transfected with *pCAG-HIVgp* and *pCMV-VSV-G-RSV-Rev* into 293T cells (on a 150-mm dish) using the FUGENE HD reagent (Roche). Twelve hours after transfection, the culture medium was replaced with fresh medium (30 ml) and the 293T cells were cultured for another 36 h. The viruses in the supernatant from each 150-mm dish were collected by ultracentrifugation and re-suspended in 1 ml of maintenance medium. For infection, mouse ES cells were dissociated and seeded onto a 35-mm dish. Eight hours after seeding, the medium was replaced with one containing the virus (1 ml per dish) and the cells were cultured for 12 h in the presence of 8 µg ml⁻¹ polybrene. After medium change, cells were further cultured for 7 days, and mRFP1-positive cells were collected by FACS. Twenty targeted clones exhibited indistinguishable abilities to differentiate into ACTH⁺ cells, and showed suppressed *Lim3* expression after Dox treatment from day 3 in SFEBq culture.

Differentiation culture for hormone-producing cells. SFEBq aggregates (10,000 cells per aggregate) were cultured in CDM medium + 400 nM SAG on days 0–10, and half the medium was changed with CDM medium on day 10. The oxygen concentration was raised to 40% from day 10 onwards. For differentiation of ACTH-lineage cells, the culture was treated with DAPT overnight on days 18–19 followed by half medium change with DAPT-free medium. For differentiation of GH- and prolactin-lineages, the aggregates were treated with BIO on days 16–18 and cultured in CDM containing 1 nM insulin and 200 ng ml⁻¹ hydrocortisone (GH) or 50 ng ml⁻¹ oestradiol (Prolactin) on days 20–30. Although these cell types showed relatively moderate differentiation efficiencies (4–6% of non-neural cells; Fig. 3k, m) as compared to corticotrophs (~35% of non-neural cells; Fig. 3c), GH⁺ and prolactin⁺ cells also appeared as cell clusters (Fig. 3l, n). In addition to these cells, the SAG-treated LCA-SFEBq aggregates also generated gonadotrophs and TSH⁺ cells when cultured in CDM-based PA6-conditioned medium during days 10–30.

Microscopy and live imaging. Confocal imaging of fixed samples was performed using the LSM710 system (Zeiss). The live imaging of Rathke's pouch formation was performed using an inverted confocal microscope combined with a full-sized CO₂/O₂ incubator (see its alignment in supplementary figure 2k of ref. 13; specially assembled by kind cooperation of Olympus), which enabled stable three-dimensional culture under long-term confocal imaging. The position of the ES cell aggregate was fixed in a drop of Matrigel, which was then immersed in culture medium on a 3.5-cm glass-bottom dish. Optical section images were obtained using a ×20 objective lens (Olympus), a spinning disk confocal system (CSU-X1, Yokogawa) and an EM-CCD camera (Andor, 512 × 512 pixels).

Immunohistochemistry, qPCR and FACS. Immunohistochemistry was performed as described previously. Antibodies against the following proteins were used at the indicated dilutions: *Lim3* (rabbit, 1:500; Chemicon), N-cadherin (mouse monoclonal, 1:2,000; BD Pharmingen), E-cadherin (rat, 1:50; TaKaRa), *Pax6* (mouse, 1:1,000; R&D, mouse monoclonal, 1:100; DSHB), *Chx10* (sheep, 1:1,000; Exalpha), *Rx* (rabbit, 1:3,000)⁹, Laminin (rat, 1:500; Chemicon; rabbit, 1:500; Abcam), PKC α /aPKC (rabbit, 1:100; Santa Cruz), CD133 (Prominin-1) (rat, 1:500; Chemicon), nestin (mouse, 1:400; BD Pharmingen; rabbit, 1:2,000; Biopad), TUJ (mouse, 1:1,000; Covance; rabbit, 1:500; Covance), neurofilament (rabbit, 1:80; Sigma), NSE (rabbit, 1:500; Chemicon), cytokeratin (mouse, 1:800; Sigma), PC2 (rabbit, 1:400; Enzo), Nkx2.1 (mouse, 1:100; Zymed; mouse, 1:500; Novocastra; rabbit, 1:2,000; Biopad), *Sox1* (rabbit, 1:200; Cell Signaling; chicken, 1:500; Chemicon), *NeuroD1* (rabbit, 1:200; Chemicon), *Pit1* (mouse, 1:100; Santa Cruz), *Sf1* (rabbit, 1:400; ABR), ACTH (mouse, 1:200; Fitzgerald), GH (rabbit, 1:400; Dako), PRL (rabbit, 1:200; AbD), α Subunit (rabbit, 1:6; NHPP), LH (rabbit, 1:400; AbD), FSH (rabbit, 1:6; NHPP), TSH (mouse, 1:50; Dako), pERK (rabbit, 1:200; Cell Signaling), GFP (rat, 1:500; Nacalai), brachyury (goat, 1:1,000; R&D) and Flk1 (rat, 1:100; BD Pharmingen). The ACTH antibody used here is specific to fully processed ACTH. The antiserum against *Pitx1* was raised in guinea pigs against a synthetic peptide (DAFKGMSLERLPEGLR+C; residues 2–18) and was affinity purified. The antiserum against *Pitx2* was raised in guinea pigs against a synthetic peptide (EKDKGQGGKNEVDGAEDPSKKKRQRQ+C; residues 16–42) and was affinity purified. The antiserum against *Tbx19* was raised in rabbit against a synthetic peptide (MSELATQKAGEGTVSRLNVESELQAGREKGDPTKE+C; residues 1–37) and was affinity purified. The monoclonal antibody for *Foxd3* was raised in rats against GST-fused mFoxD3 protein (residues 232–465) and its culture sup was used at 1/10 dilution. DAPI was used for counterstaining the nuclei (Molecular Probes). The immunostaining specificity of each antibody was confirmed by immunostaining with the appropriate embryonic tissues as a positive control under the same conditions.

For FACS analysis, cells were counted with a FACSAria (Becton Dickinson), and the data were analysed with the FACSDiva software (Becton Dickinson).

qPCR was performed with 8 aggregates per sample using the 7500 Fast Real Time PCR System (Applied Biosystems) and the data were normalized to

the GAPDH expression. Primers used were as follows: *GAPDH*, forward 5'-TGACCACAGTCCATGCCATC-3', reverse 5'-GACGGACACATTGGGGG TAG-3'; *Pitx2*, forward 5'-CGTGTGGACCAACCTTACG-3', reverse 5'-AAGC CATTCTTGCACAGCTC-3'; *Wnt3a*, forward 5'-GAACCGTCACAACAATG AGG-3', reverse 5'-CTTCACAGCTGCCAGATAGC-3'; *Wnt4*, forward 5'-GG CCTTTGTATACGCCATCT-3', reverse 5'-CACAGCCACACTTCTCCAGT-3'; *Wnt5a*, forward 5'-ACGCTAGAGAAAGGGAACGA-3', reverse 5'-ATACTGT CCTACGGCCTGCT-3'; *Lim3*, forward 5'-ACAAGGACAGCATCCAGGAG-3', reverse 5'-GCTGCTGTACAGGCCATTAG-3'; *Fgf8*, forward 5'-GGGAAGCTA ATTGCCAAGAG-3', reverse 5'-TGTACCAGCCCTCGTACTTG-3'; *Fgf10*, forward 5'-TGCTCTTTTTGGTGTCTTCG-3', reverse 5'-GGAGGAAGAAGA GCAGTTGG-3'; *Bmp4*, forward 5'-GCTGGAATGATTGGATTGTG-3', reverse 5'-CATGGTTGGTTGAGTTGAGG-3'; *Bmp2*, forward 5'-CTCAAGTCCAGCT GCAAGAG-3', reverse 5'-ACTCCCCATGGCAGTAAAG-3'; *Rx*, forward 5'-GTTCCGGGTCCAGGTATGGT-3', reverse 5'-GAGAGGAGGGGAGAATC CTG-3'; *NeuroD1*, forward 5'-TTGAAGCCATGAATGCAGAG-3', reverse 5'-CCTCCTCTCCTCTAGATCCTCA-3'; *Pit1*, forward 5'-ACCACAGTGC CGCTGAGT-3', reverse 5'-GTAATGAAGTCCTGTCGTGTG-3'; *Tbx19*, forward 5'-CTGTCTGTACCCACAGCAA-3', reverse 5'-TCCACAGGCAC GGATAGG-3'; *ACTH/Pomc*, forward 5'-GGCCTTTCCCCTAGAGTTCA-3', reverse 5'-GACCTGCTCCAAGCCTAATG-3'; *Fgf4*, forward 5'-CAGAGGCCT TTGGTATGGAT-3', reverse 5'-AGGTCTGCCAAATCCTTGTC-3'; *Fgf2b*, forward 5'-AGTAAATACGGGCCTGATGG-3', reverse 5'-CCATCTCCGT CACATTGAAC-3' and brachyury, forward 5'-CTGGGAGCTCAGTTCT TTCG-3', reverse 5'-CCCTTCATACATCGGAGAA-3'.

Statistical analyses. Statistical significance was tested using the Prism 4 program (GraphPad) as follows: two-group comparison by Student's *t*-test and multiple-group comparison by the one-way ANOVA test with a post-hoc Bonferroni's (among each pair) or Dunnett's test (versus control).

In vitro analysis of ACTH release by CRH loading. Eight SAG+DAPT-treated aggregates were collected on day 23 in a 1.5-ml Eppendorf tube, rinsed with HBSS(+), and pre-incubated in 500 μ l HBSS at 37 °C for 10 min. CRH was then added to the given concentrations. The supernatant was collected after 10-min incubation at 37 °C and subjected to ELISA using the ACTH ELISA kit (MD Bioproducts). In contrast to ACTH, no enhanced secretion of GH, FSH, LH, TSH or prolactin was significantly

induced by CRH (not shown). In addition, although the basal secretion (without releasing factor treatment) of ACTH was reproducibly observed (Fig. 4), the basal secretion levels of mouse GH, PRL, FSH, LH and TSH from SAG+DAPT-treated aggregates were below the detection limits of ELISA (SRL).

Transplantation of ES-cell-derived pituitary tissues. All animal experiments were performed by following the institutional (RIKEN) guidelines for animal studies. Surgical hypophysectomy of mice (ICR background; 8-week-old males) was performed by the transaural approach described previously³³. Briefly, mice were anaesthetized with intraperitoneal (i.p.) injection of pentobarbital (40 mg kg⁻¹), and the pituitary tissues were aspirated from the sella turcica using a needle (KN-390 needle, Natsume Seisakusyo) set to 1 ml syringe containing 0.2 ml saline, following its perforation via the auditory meatus. Seven days later (9-week-old), basal and CRH-stimulated (2 μ g kg⁻¹, i.p.) blood levels of ACTH were analysed to confirm the hypopituitarism status. One week later (10-week-old), the hypopituitary mice were anaesthetized and injected with ~40 SAG+DAPT-treated or control day-21 aggregates (in 100 μ l phosphate-buffered saline) into the unilateral kidney under the capsule using a 23G needle syringe (the kidney was exposed by skin/muscle/peritoneum incision via the dorso-lateral approach). Prior to the operation, the mice were injected with dexamethasone (0.3 mg, one shot, intramuscularly; this glucocorticoid supplementation was essential for good post-operational survival of hypopituitary mice, which are very weak against physical stress; for one-shot supplementation, dexamethasone worked better than very short-life steroids such as hydrocortisone). Following the operation, 100 mg ampicillin was administered intramuscularly. Seven days later (11 week-old), the grafted mice were subjected to a CRH-loading test (2 μ g kg⁻¹, i.p.) and blood sampling was done 2 h after the CRH injection. The 'basal' secretion levels of ACTH and corticosterone (without CRH loading) were examined at 20:00 in a stress-free environment. It remains to be clarified at present whether this moderate ACTH release from the grafted S+D+ aggregates represented solely their spontaneous basal secretion or also included some secretion weakly stimulated by peripheral CRH as well as vasopressin, which is also known to evoke some ACTH release from corticotrophs.

The spontaneous-locomotor tests (11-week-old male mice) were performed under a stress-free condition in the mice's home cages using an infrared 24-h monitoring system (MDC-W02; BrainScienceIdea) and a running-wheel device (ENV-044; MedAssociates), respectively.

Mutations causing syndromic autism define an axis of synaptic pathophysiology

Benjamin D. Auerbach¹, Emily K. Osterweil¹ & Mark F. Bear¹

Tuberous sclerosis complex and fragile X syndrome are genetic diseases characterized by intellectual disability and autism. Because both syndromes are caused by mutations in genes that regulate protein synthesis in neurons, it has been hypothesized that excessive protein synthesis is one core pathophysiological mechanism of intellectual disability and autism. Using electrophysiological and biochemical assays of neuronal protein synthesis in the hippocampus of *Tsc2*^{+/-} and *Fmr1*^{-/-} mice, here we show that synaptic dysfunction caused by these mutations actually falls at opposite ends of a physiological spectrum. Synaptic, biochemical and cognitive defects in these mutants are corrected by treatments that modulate metabotropic glutamate receptor 5 in opposite directions, and deficits in the mutants disappear when the mice are bred to carry both mutations. Thus, normal synaptic plasticity and cognition occur within an optimal range of metabotropic glutamate-receptor-mediated protein synthesis, and deviations in either direction can lead to shared behavioural impairments.

More than 1% of the human population has an autism spectrum disorder (ASD), and it has been estimated that over 50% of those with autism also have intellectual disability¹. In most cases, the cause is unknown. However, genetically defined syndromes with increased prevalence of autism and intellectual disability offer an opportunity to understand the brain pathophysiology that manifests as ASD and intellectual disability, and this knowledge can suggest potential therapies. A case in point is fragile X syndrome (FXS), caused by silencing of the *FMR1* gene and loss of the protein product, FMRP. Studies of the *Fmr1* knockout (-/-) mouse revealed that in the absence of FMRP, protein synthesis is increased downstream of metabotropic glutamate receptor 5 (mGluR5). Diverse mutant phenotypes in fragile X animal models have been corrected by genetic or pharmacological inhibition of mGluR5, and preliminary human clinical trials using drugs that inhibit mGluR5 have shown promise². Because several other syndromic forms of ASD and intellectual disability are associated with mutations of genes that regulate messenger RNA (mRNA) translation at synapses, it has been hypothesized that altered synaptic protein synthesis might contribute generally to the autistic phenotype, including intellectual disability³. The aim of the current study was to test the hypothesis that a mutation responsible for another genetic syndrome associated with ASD and intellectual disability—tuberous sclerosis complex (TSC)—produces abnormalities in synaptic protein synthesis and plasticity similar to fragile X. If this were the case, treatments developed for one disorder might be beneficial for the other, and possibly for autism and intellectual disability more broadly.

TSC mutations affect synaptic function

The choice of TSC was guided by several considerations. Like FXS, (1) TSC is a single-gene disorder with core symptoms of ASD and intellectual disability, (2) the affected gene(s) lie in a signalling pathway that couples cell surface receptors to mRNA translation, (3) there are well validated mouse models of the disease and (4) some mutant phenotypes in these mouse models have responded to pharmacological treatments that affect protein synthesis^{4–6}. The disease is caused by heterozygous mutations in the genes encoding TSC1 or TSC2 proteins that together

form the TSC1/2 complex. TSC1/2 acts to inhibit Rheb, a Ras family GTPase with high specificity for mTOR within a protein complex called mTORC1. Rheb activation of mTORC1 can stimulate mRNA translation and cell growth, and excessive mTORC1 activation is believed to be pathogenic in TSC⁷. TSC is characterized by the growth of hamartomas that are believed to result from inactivation of the functional allele within the tumour cells^{8,9}. Although some neurological manifestations of TSC are thought to be related to tumour growth in the cerebral cortex, others including cognitive impairment and autism have been proposed to result from abnormal signalling at synapses¹⁰. Consistent with this idea, mice engineered to carry heterozygous loss-of-function mutations in *Tsc1* or *Tsc2* have been shown to have hippocampus-dependent learning and memory deficits without having tumours in the brain or seizures^{4,11}. Here we chose the *Tsc2*^{+/-} mouse model because TSC2 mutations are more common and produce a more severe phenotype in humans¹², and this animal model is in widespread use^{4,13–15}. Of particular significance, postnatal treatment of *Tsc2*^{+/-} mice with the mTORC1 inhibitor rapamycin was previously shown to ameliorate hippocampal memory impairments, suggesting the exciting possibility that some aspects of TSC, like FXS, might be amenable to drug therapy⁴.

A prominent hypothesis is that synaptic dysfunction in TSC relates to increased protein synthesis in response to elevated mTORC1 activity¹⁶. Signalling by mTORC1 has been suggested to contribute to the coupling of mGluR5 to protein synthesis and, although still controversial, it has been proposed that elevated mTOR activity might also be a cause of elevated protein synthesis in the *Fmr1*^{-/-} mouse¹⁷. A sensitive electrophysiological read-out of local mRNA translation in response to mGluR5 activation is long-term synaptic depression (LTD) in area CA1 of the hippocampus^{18,19}. Indeed, it was exaggerated LTD in the *Fmr1*^{-/-} mouse that led to the mGluR theory of FXS^{20,21}. Therefore, to test the hypothesis of a shared pathophysiology between TSC and FXS, we first examined mGluR-LTD in the hippocampus of male *Tsc2*^{+/-} mice.

Protein synthesis and mGluR-LTD in *Tsc2*^{+/-} mice

LTD was induced by activation of group 1 (Gp1) mGluRs (mGluR 1 and 5) with the selective agonist DHPG ((R,S)-3,5-dihydroxyphenylglycine)

¹Howard Hughes Medical Institute, The Picower Institute for Learning and Memory, Department of Brain and Cognitive Sciences, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA.

in hippocampal slices¹⁹. Unexpectedly, we discovered that DHPG-induced LTD was deficient rather than enhanced in the hippocampus of *Tsc2*^{+/-} mice, compared with wild-type (WT) controls (Fig. 1a). A similar deficit was observed when mGluR-LTD was induced by patterned electrical stimulation of Schaffer collateral synapses (Fig. 1b). In agreement with a previous report⁴, basal synaptic transmission in CA1 appeared normal in the *Tsc2*^{+/-} mice, indicating that the impairment in mGluR-LTD is not due to general disruption of synaptic function (Supplementary Fig. 1). Moreover, there was no difference in the magnitude of the NMDA (*N*-methyl-D-aspartate)-receptor-dependent form of LTD between WT and *Tsc2*^{+/-} mice (Fig. 1c), demonstrating that the deficit is specific to mGluR-LTD, as these same synapses are able to undergo activity-induced depression by a different mechanism. To test the possibility of a general disruption in Gp 1 mGluR function, we examined DHPG-induced phosphorylation of extracellular signal-regulated kinase 1/2 (ERK1/2), a common measure of Gp1 mGluR signalling and a critical step for mGluR-mediated protein synthesis and LTD^{22,23}. Basal ERK1/2 phosphorylation and DHPG-induced increases in ERK1/2 phosphorylation are unaltered in *Tsc2*^{+/-} mice (Fig. 1d). These results suggest that the deficit in mGluR-LTD seen in the *Tsc2*^{+/-} hippocampus is not due to a global dysregulation of synaptic function or Gp 1 mGluR signalling.

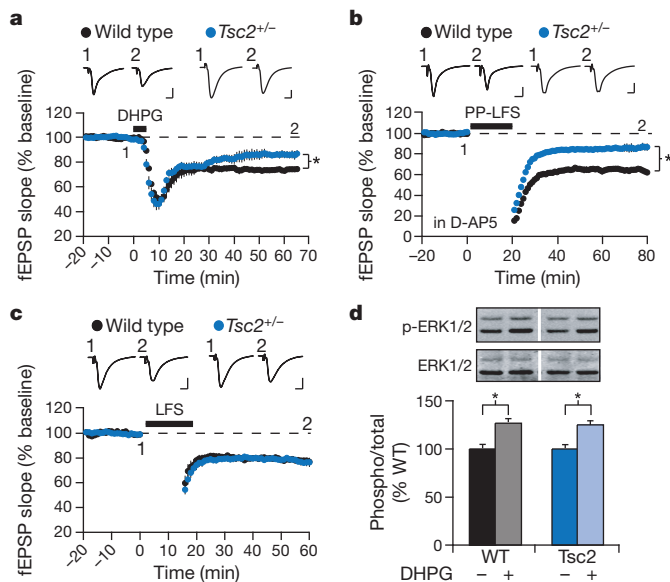


Figure 1 | *Tsc2*^{+/-} mice have a specific deficit in mGluR-LTD. **a**, DHPG induces significantly less LTD in slices from *Tsc2*^{+/-} mice versus littermate WT mice (WT: 74.3 ± 1.4%, *n* = 5 animals, 10 slices; *Tsc2*^{+/-}: 86.3 ± 3.1%, *n* = 6 animals, 12 slices; **P* = 0.004). In this and all subsequent electrophysiology figures, representative field potential traces (average of 10 sweeps) were taken at times indicated by numerals and scale bars equal to 0.5 mV, 5 ms, unless stated otherwise. Error bars, s.e.m. **b**, Synaptically induced mGluR-LTD, elicited by delivering pairs of pulses (50 ms inter-stimulus interval) at 1 Hz for 20 min (paired-pulse low-frequency synaptic stimulation (PP-LFS), 1,200 pulses) in the presence of the NMDA receptor antagonist D-(-)-2-amino-5-phosphonopentanoic acid (D-AP5, 50 μM), is also deficient in slices from *Tsc2*^{+/-} mice (WT: 65.1 ± 2.1%, *n* = 3 animals, 9 slices; *Tsc2*^{+/-}: 85.0 ± 2.5%, *n* = 4 animals, 11 slices; **P* = 0.003). **c**, The magnitude of NMDA receptor-dependent LTD evoked by low-frequency stimulation (LFS, 900 pulses at 1 Hz) does not differ between genotypes (WT: 79.8 ± 1.6%, *n* = 4 animals, 6 slices; *Tsc2*^{+/-}: 79.4 ± 1.9%, *n* = 6 animals, 6 slices; *P* = 0.610). **d**, Hippocampal slices were stimulated with 50 μM DHPG for 5 min, and ERK1/2 activation (phosphorylation) assessed by immunoblot (normalized WT: 100.0 ± 6.1%; WT DHPG: 119.6 ± 5.5%; *Tsc2*^{+/-}: 97.5 ± 5.6%; *Tsc2*^{+/-} DHPG: 116.2 ± 3.9%; ANOVA: genotype *P* = 0.623, treatment **P* = 0.0008, genotype × treatment *P* = 0.923; *n* = 9 animals). Results reveal that DHPG significantly increases ERK1/2 activation in both WT (**P* = 0.040) and *Tsc2*^{+/-} (**P* = 0.003). Error bars, s.e.m.

mGluR-LTD in area CA1 of the hippocampus is expressed by two independent mechanisms: reduced probability of presynaptic glutamate release^{24–26} and reduced expression of postsynaptic AMPA receptors^{25,27}. In WT animals, the postsynaptic modification is known to require immediate translation of mRNAs available in the dendrites of hippocampal pyramidal neurons^{18,28}. Accordingly, we found that LTD in WT mice at the age range examined (postnatal day (P) 25–35) is reliably reduced by the protein synthesis inhibitor cycloheximide (60 μM; Fig. 2a). The presynaptic component of LTD was monitored by measuring paired-pulse facilitation, which showed a persistent increase following DHPG that reflects reduced probability of glutamate released at the presynaptic terminal^{24–26}. Changes in paired-pulse facilitation were not inhibited by cycloheximide (Supplementary Fig. 2 and Fig. 2c), suggesting that residual LTD in the presence of the drug is expressed presynaptically. Although LTD was reduced in *Tsc2*^{+/-} mice, the persistent paired-pulse facilitation change after DHPG was no different than in WT, suggesting a deficient postsynaptic modification (Supplementary Fig. 2 and Fig. 2c). Indeed, unlike WT, cycloheximide treatment had no effect on LTD in the *Tsc2*^{+/-} animals (Fig. 2b). These data suggest a selective loss of the protein-synthesis-dependent component of LTD in the mutant mice.

These electrophysiological results in the *Tsc2*^{+/-} hippocampus stand in stark contrast to the *Fmr1*^{-/-} mouse in which mGluR-LTD is exaggerated²⁰. In the fragile X mouse model, increased LTD correlates with an increased rate of basal mRNA translation downstream of mGluR5. Therefore we were compelled to examine protein synthesis in hippocampal slices from the *Tsc2*^{+/-} mouse as previously described for the *Fmr1*^{-/-} mouse²³. Consistent with the mGluR-LTD findings, we found a small but significant decrease in [³⁵S]methionine/cysteine incorporation into protein under basal conditions in the hippocampus of *Tsc2*^{+/-} mice (Fig. 2d). This finding suggested the possibility that protein(s) required for mGluR-LTD are deficiently translated in the hippocampus of *Tsc2*^{+/-} mice. To test this idea we examined levels of Arc, a plasticity-related protein that is rapidly synthesized in response to Gp 1 mGluR activation and is required for mGluR-LTD^{29,30}. Interestingly, we found that Arc expression is decreased in *Tsc2*^{+/-} hippocampal slices (Fig. 2e). To determine whether this decrease was due to diminished translation, we measured the amount of newly synthesized Arc in *Tsc2*^{+/-} slices by performing immunoprecipitation experiments on metabolically labelled slices (see Methods)²³. Examination of the ³⁵S-incorporated fraction revealed a significant reduction in Arc translation in the hippocampus of *Tsc2*^{+/-} mice (Fig. 2f). Control immunoprecipitations using non-immune IgG confirmed that our measurements were specific for Arc (Supplementary Fig. 4). These results suggest that mGluR-LTD is deficient in the *Tsc2*^{+/-} hippocampus because of a decrease in the translation of the proteins required to stabilize LTD, including Arc.

LTD deficit caused by excess mTOR activity

As in the human disease, the germline mutation in *Tsc2* can have myriad secondary consequences on neural development that could contribute to the observed LTD and protein synthesis phenotypes. To test the hypothesis that the deficient mGluR-LTD seen in *Tsc2*^{+/-} mice is a specific consequence of unregulated mTOR activity, we examined the effects of the mTORC1 inhibitor rapamycin. We found that acute rapamycin treatment (20 nM) restored mGluR-LTD in the *Tsc2*^{+/-} mice to WT levels (Fig. 2g), whereas this same treatment had no effect on mGluR-LTD in slices from WT mice (Supplementary Fig. 3). This rescue is due specifically to the recovery of the protein-synthesis-dependent component of LTD, as the effect of rapamycin in *Tsc2*^{+/-} mice was eliminated in the presence of cycloheximide (Fig. 2h). The same rapamycin treatment also restored basal protein synthesis rates in *Tsc2*^{+/-} hippocampal slices back to WT levels (Fig. 2i). The simple model that best fits the data is that unregulated mTOR activity caused by the *Tsc2*^{+/-} mutation suppresses the protein synthesis that is required for mGluR-LTD (Fig. 3a).

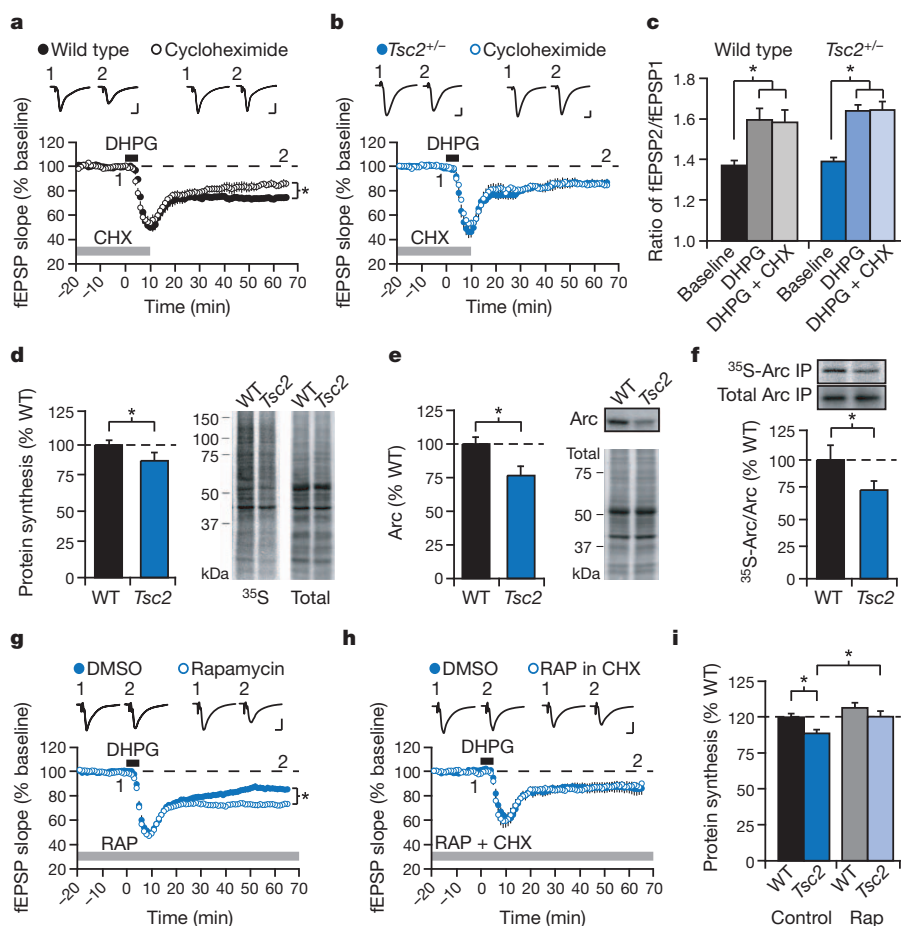


Figure 2 | Excessive mTOR activity suppresses the protein-synthesis-dependent component of mGluR-LTD. **a**, LTD is significantly attenuated by pretreatment with the protein synthesis inhibitor cycloheximide (CHX, 60 μ M, grey bar) in slices from WT animals (control: $74.3 \pm 1.4\%$, $n = 5$ animals, 10 slices; CHX: $85.2 \pm 2.8\%$, $n = 4$ animals, 7 slices; $*P = 0.014$). **b**, CHX treatment has no effect on LTD in slices from *Tsc2*^{+/-} mice (control: $86.3 \pm 3.1\%$, $n = 6$ animals, 12 slices; CHX: $85.3 \pm 3.2\%$, $n = 4$ animals, 7 slices, $P = 0.796$). ANOVA: genotype $*P = 0.041$, treatment $P = 0.089$, genotype \times treatment $*P = 0.045$. **c**, Presynaptic LTD is not affected by genotype or CHX (see also Supplementary Fig. 2). DHPG significantly increased paired-pulse facilitation in slices from both WT and *Tsc2*^{+/-} mice (paired-pulse facilitation with a 50 ms inter-stimulus interval: WT baseline: 1.37 ± 0.02 , WT DHPG: 1.59 ± 0.06 , $n = 5$ animals, 9 slices, $*P = 0.003$; *Tsc2*^{+/-} baseline: 1.39 ± 0.02 , *Tsc2*^{+/-} DHPG: 1.64 ± 0.03 , $n = 5$ animals, 9 slices, $*P = 0.001$) and this effect was not blocked by CHX (WT DHPG + CHX: 1.58 ± 0.06 , $n = 7$ animals, 11 slices, $P = 0.89$; *Tsc2*^{+/-} DHPG + CHX: 1.64 ± 0.04 , $n = 6$ animals, 7 slices, $P = 0.94$). **d**, Metabolic labelling of hippocampal slices reveals a significant reduction of basal protein synthesis in *Tsc2*^{+/-} mice (WT: $100.0 \pm 3.1\%$, *Tsc2*^{+/-}: $88.2 \pm 3.3\%$, $n = 13$ animals; $*P = 0.043$). Differences in protein synthesis are exemplified by representative autoradiograph and total protein stain of the same membrane. **e**, Immunoblotting experiments show that Arc expression is significantly reduced in *Tsc2*^{+/-} hippocampal slices (WT:

$100.0 \pm 4.7\%$, *Tsc2*^{+/-}: $76.6 \pm 6.4\%$, $n = 12$ animals; $*P = 0.005$). **f**, Arc translation was measured by metabolic labelling of hippocampal slices, followed by immunoprecipitation of Arc. Comparison of the ratios of ³⁵S-incorporated-to-total Arc reveals a significant reduction in Arc translation in the *Tsc2*^{+/-} hippocampus (WT: $100.0 \pm 11.5\%$, *Tsc2*^{+/-}: $74.7 \pm 6.8\%$, $n = 19$ animals; $*P = 0.049$). **g**, Pretreatment of slices with the mTORC1 inhibitor rapamycin (RAP, 20 nM, grey bar) significantly enhances DHPG-induced LTD in slices from *Tsc2*^{+/-} mice (dimethylsulphoxide (DMSO): $85.7 \pm 2.1\%$, $n = 8$ animals, 17 slices; RAP: $72.9 \pm 1.8\%$, $n = 7$ animals, 18 slices; $*P = 0.002$). **h**, The rescue by rapamycin of DHPG-induced LTD in *Tsc2*^{+/-} mice is prevented by the protein synthesis inhibitor cycloheximide (DMSO): $87.1 \pm 4.7\%$, $n = 6$ animals, 10 slices; RAP: $88.1 \pm 2.4\%$, $n = 7$ animals, 9 slices; $P = 0.796$). ANOVA: rapamycin treatment $*P = 0.043$, cycloheximide treatment $*P = 0.004$, rapamycin \times cycloheximide $*P = 0.018$. **i**, Metabolic labelling experiments show that rapamycin (20 nM) normalizes protein synthesis in the *Tsc2*^{+/-} hippocampus to WT levels (WT DMSO: $100.0 \pm 2.5\%$, WT RAP: $106.5 \pm 3.6\%$, *Tsc2*^{+/-} DMSO: $88.8 \pm 2.6\%$, *Tsc2*^{+/-} RAP: $100.4 \pm 3.9\%$; ANOVA: genotype $*P = 0.008$, treatment $*P = 0.006$, genotype \times treatment $P = 0.430$; t -test: WT compared with *Tsc2*^{+/-} DMSO $*P = 0.003$; WT compared with *Tsc2*^{+/-} RAP $P = 0.344$; *Tsc2*^{+/-} DMSO compared with RAP $*P = 0.037$; $n = 22$ animals). Error bars, s.e.m.

Effect of mGluR5-positive allosteric modulation

In the *Fmr1*^{-/-} model of FXS, excessive mGluR-LTD and hippocampal protein synthesis can be corrected by reducing signalling by mGluR5^{23,31}. We therefore wondered if the opposite approach of potentiating mGluR5 signalling with a positive allosteric modulator (PAM) could be beneficial in this model of TSC (Fig. 3a). PAMs are compounds that do not activate mGluR5 directly but act on an allosteric site to potentiate physiological activation of the receptor³². Indeed, we found that pretreatment of hippocampal slices with the mGluR5 PAM 3-cyano-N-(1,3-diphenyl-1H-pyrazol-5-yl)benzamide (CDPPB³³) restored the magnitude of mGluR-LTD in *Tsc2*^{+/-} mice to

WT levels (Fig. 3b). The rescue of LTD appears to be due specifically to recovery of the protein-synthesis-dependent component because the effect of CDPPB was completely eliminated by cycloheximide (Fig. 3c). Consistent with this conclusion, CDPPB treatment also restored basal protein synthesis levels (Fig. 3d) and rescued the deficit in Arc synthesis in the *Tsc2*^{+/-} mice (Fig. 3e). Thus, allosteric augmentation of mGluR5 signalling can overcome the inhibitory effect of unregulated mTOR activity on the synaptic protein synthesis that supports LTD.

In an important recent study, cognitive impairments in the *Tsc2*^{+/-} mice were shown to be significantly improved by treating the animals with the mTORC1 inhibitor rapamycin⁴. In light of our

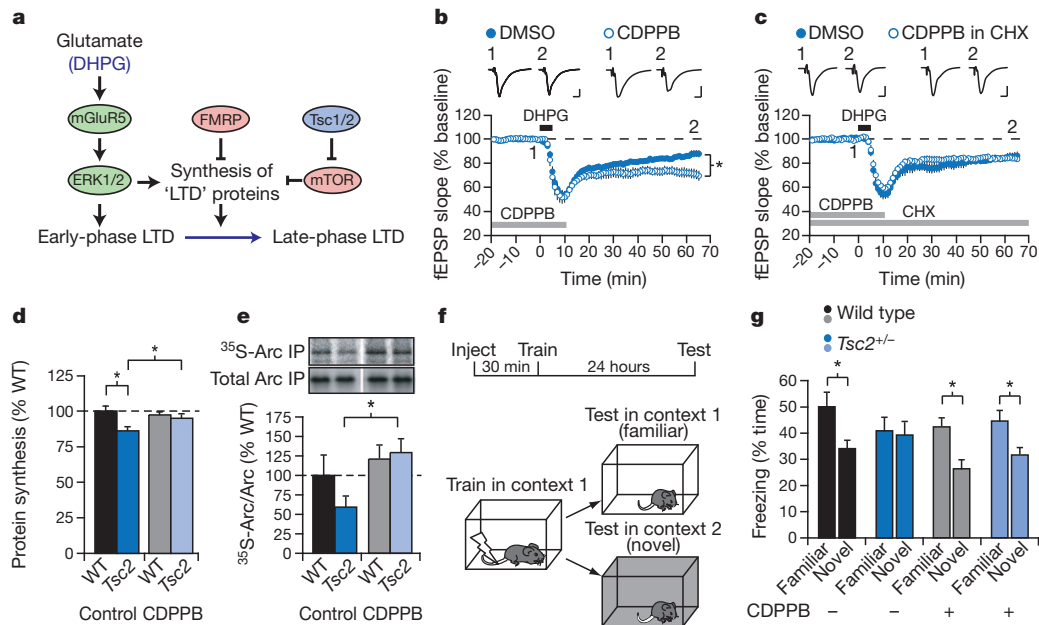


Figure 3 | Positive modulation of mGluR5 reverses synaptic and behavioural deficits in *Tsc2*^{+/-} mice. **a**, Model to account for effects of *Tsc2*^{+/-} and *Fmr1*^{-/-} mutations on mGluR5- and protein-synthesis-dependent LTD. This model predicts that the impairment in *Tsc2*^{+/-} mice can be overcome either by inhibiting mTOR with rapamycin or by augmenting mGluR5 signalling with an mGluR5 PAM. **b**, Consistent with the model, pretreatment of slices from *Tsc2*^{+/-} mice with CDPBPB (10 μ M, grey bar) significantly enhances DHPG-induced LTD (DMSO: $86.4 \pm 2.5\%$, $n = 8$ animals, 13 slices; CDPBPB: $71.7 \pm 3.9\%$, $n = 7$ animals, 12 slices; $*P < 0.001$). **c**, CDPBPB treatment fails to enhance DHPG-induced LTD in *Tsc2*^{+/-} mice when co-applied with the protein synthesis inhibitor cycloheximide (DMSO: $89.0 \pm 4.4\%$, $n = 8$ animals, 10 slices; CDPBPB: $83.9 \pm 2.1\%$, $n = 7$ animals, 9 slices; $P = 0.64$). ANOVA: CDPBPB treatment $*P = 0.008$, CHX treatment $P = 0.087$, CDPBPB \times CHX $*P = 0.034$. **d**, CDPBPB (10 μ M) restores protein synthesis in the *Tsc2*^{+/-} hippocampus to WT levels (WT DMSO: $100.0 \pm 3.2\%$, WT CDPBPB: $97.2 \pm 1.9\%$, *Tsc2*^{+/-} DMSO: $86.1 \pm 2.7\%$, *Tsc2*^{+/-} CDPBPB: $94.9 \pm 3.0\%$; ANOVA: genotype $*P = 0.006$, treatment $P = 0.275$, genotype \times treatment $*P = 0.041$; *t*-test: WT compared with

Tsc2^{+/-} DMSO $*P = 0.012$; WT compared with *Tsc2*^{+/-} CDPBPB $P = 0.538$; *Tsc2*^{+/-} DMSO compared with CDPBPB $*P = 0.049$; $n = 17$ animals). **e**, CDPBPB exposure significantly increases Arc translation in the *Tsc2*^{+/-} hippocampus (WT DMSO $100.0 \pm 28.2\%$, WT CDPBPB $121.0 \pm 21.2\%$, *Tsc2*^{+/-} DMSO $59.2 \pm 7.0\%$, *Tsc2*^{+/-} CDPBPB $129.4 \pm 20.3\%$; ANOVA genotype $P = 0.554$, treatment $*P = 0.009$, genotype \times treatment $P = 0.114$; *t*-test: *Tsc2*^{+/-} DMSO compared with CDPBPB $*P = 0.026$; $n = 6$ animals). Error bars, s.e.m. **f**, Experimental design of context discrimination task. **g**, WT mice display intact memory by freezing more in the familiar context than the novel context (black bars; familiar: $50 \pm 7.7\%$, $n = 12$; novel: $34.1 \pm 3.2\%$, $n = 14$; $*P = 0.003$). A single injection of CDPBPB (10 mg kg⁻¹, intraperitoneal) 30 min before training has no effect on WT context discrimination (familiar: $42.3 \pm 3.7\%$, $n = 12$; novel: $26.4 \pm 3.6\%$, $n = 12$; $*P = 0.005$). Control *Tsc2*^{+/-} mice display an impairment in context discrimination (blue bars; familiar: $40.9 \pm 5.3\%$, $n = 11$; novel: $39.3 \pm 5.2\%$, $n = 14$; $P = 0.501$), but this deficit is corrected by a single injection of CDPBPB (familiar: $44.5 \pm 4.3\%$, $n = 11$; novel: $31.6 \pm 3\%$, $n = 12$; $*P = 0.034$). Error bars, s.e.m.

electrophysiological and biochemical findings, we wondered if a similar amelioration would be observed with the mGluR5 PAM. A robust phenotype was reported to be an impairment in the ability of the *Tsc2*^{+/-} mice to distinguish between familiar and novel contexts in a fear conditioning task. Advantages of this task are that the learning occurs in one trial, making it amenable to acute drug treatment, and the memory is hippocampus dependent³⁴. Although a requirement for CA1 LTD per se has not been established, contextual fear discrimination does depend on both mGluR5³⁵ and new protein synthesis at the time of training³⁶. In this assay, mice are first exposed to a distinctive context in which they receive an aversive foot shock. The next day, context discrimination is tested by dividing the animals into two groups: one is placed in the familiar context associated with the shock; the other is placed in a novel context (Fig. 3f). Context discrimination is assessed by measuring the time the animals express fear by freezing in each context. Although the WT mice clearly discriminate between contexts, the *Tsc2*^{+/-} mice do not⁴ (Fig. 3g). To test the effect of augmenting mGluR5 signalling, mice from both genotypes were injected intraperitoneally with CDPBPB (10 mg kg⁻¹) 30 min before training. Although this treatment had no effect in the WT mice, it was sufficient to correct the deficit in context discrimination observed in the *Tsc2*^{+/-} mice. These results show that augmentation of mGluR5 signalling is beneficial at the behavioural level in *Tsc2*^{+/-} mice and that disrupted mGluR5 function may be relevant to cognitive impairments associated with TSC.

Fmr1^{-/-} and *Tsc2*^{+/-} mutations cancel each other

Contrary to our initial hypothesis, we found that mutations causing FXS and TSC, two disorders associated with autism and intellectual disability, show mirror symmetrical alterations in protein-synthesis-dependent LTD and have beneficial responses to treatments that modulate mGluR5 in opposite directions (Fig. 4a). These findings raised the intriguing possibility that these two mutations could cancel one another on this functional axis. To test this hypothesis, we introduced an *Fmr1* deletion into the *Tsc2*^{+/-} background by crossing *Tsc2*^{+/-} males with *Fmr1*^{+/-} females (Fig. 4b). This approach also enabled us to compare directly with WT the effects of the *Tsc2*^{+/-} and *Fmr1*^{-/-} mutations in littermates reared under identical conditions. As expected, mGluR-LTD was diminished in *Tsc2*^{+/-} mice and excessive in the *Fmr1*^{-/-} mice, compared with WT (Fig. 4c,d). However, mice harbouring both mutations showed mGluR-LTD that was indistinguishable from WT (Fig. 4c,d).

Although *Tsc2*^{+/-} and *Fmr1*^{-/-} mutations cause opposite alterations in mGluR-LTD and protein synthesis, the human disorders they are associated with have similar neurological and cognitive phenotypes. Might opposite deviations in synaptic function lead to shared cognitive impairments? To examine this question, we compared context discrimination in the *Tsc2*^{+/-} and *Fmr1*^{-/-} mice and discovered that indeed they do share a deficit in this measure of memory (Fig. 4e). Remarkably, instead of being exacerbated, this memory deficit was erased in the double mutants (Fig. 4e). These results suggest that the

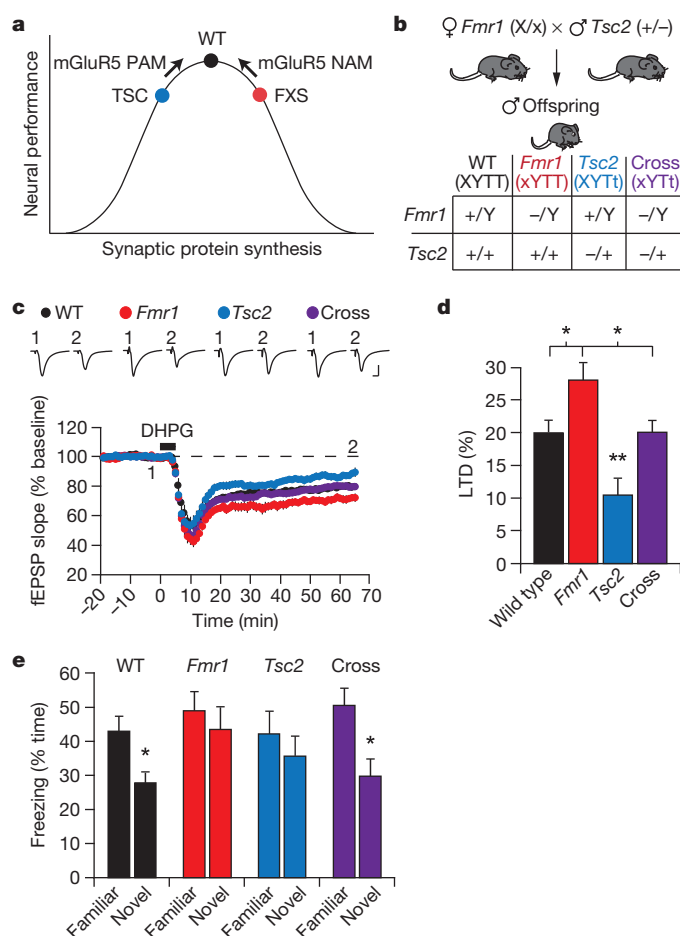


Figure 4 | Genetic cross of *Tsc2*^{+/-} and *Fmr1*^{-/-} mice rescues synaptic and behavioural impairments present in both single mutants. **a**, The data suggest that optimal synaptic function requires a narrow and tightly regulated level of synaptic protein synthesis and that deviations in either direction can impair function^{3,43}. TSC and FXS fall on different ends of this spectrum and respond to opposite alterations of mGluR5 signalling. These results raise the possibility that introducing both mutations to a mouse may normalize aspects of neural function. **b**, Genetic rescue strategy. Heterozygous *Tsc2* male mice (*Tsc2*^{+/-}) were bred with heterozygous *Fmr1* females (*Fmr1*^{+/-}) to obtain male offspring of four genotypes: WT (*Tsc2*^{+/+}, *Fmr1*^{+/+}), *Fmr1* KO (*Tsc2*^{+/+}, *Fmr1*^{-/-}), *Tsc2* Het (*Tsc2*^{+/-}, *Fmr1*^{+/+}) and cross (*Tsc2*^{+/-}, *Fmr1*^{-/-}). **c**, DHPG-induced LTD is significantly decreased in slices from *Tsc2*^{+/-} mice (**P* = 0.002) and significantly increased in slices from *Fmr1*^{-/-} mice (**P* = 0.017), compared with WT slices. DHPG-LTD in slices from *Tsc2*^{+/-} × *Fmr1*^{-/-} mice is comparable in magnitude to WT slices (*P* = 0.558). (WT: 78.9 ± 2.1%, *n* = 7 animals, 17 slices; *Fmr1*: 71.2 ± 2.7%, *n* = 7 animals, 21 slices; *Tsc2*: 89.5 ± 2.6%, *n* = 7 animals, 15 slices; cross: 77.4 ± 1.8%, *n* = 9 animals, 19 slices). **d**, Summary of LTD data. Bar graphs, percentage decrease from baseline in fEPSP (average of last 5 min of recording ± s.e.m.); **P* < 0.05, ***P* < 0.01. **e**, Both mutations cause a deficit in context discrimination that is rescued in the double mutant. WT mice (familiar: 42.9 ± 4.6%, *n* = 11; novel: 27.8 ± 3.4%, *n* = 12; **P* = 0.024), *Fmr1*^{-/-} mice (familiar: 49.0 ± 5.6%, *n* = 11; novel: 43.5 ± 6.7%, *n* = 12; *P* = 0.483), *Tsc2*^{+/-} mice (familiar: 42.1 ± 6.8%, *n* = 12; novel: 35.6 ± 6.0%, *n* = 12; *P* = 0.395) and *Tsc2*^{+/-} × *Fmr1*^{-/-} mice (familiar: 50.5 ± 5.2%, *n* = 11; novel: 29.8 ± 5.2%, *n* = 11; **P* = 0.011). Error bars, s.e.m.

opposing synaptic deviations seen in *Tsc2*^{+/-} and *Fmr1*^{-/-} mice may manifest similarly at the behavioural level, as introducing both mutations not only reverses the disruptions of synaptic plasticity but rescues this memory impairment as well.

Discussion

LTD and protein synthesis downstream of mGluR5 have attracted attention in the context of several diseases, most notably FXS²⁷.

Fragile X is caused by the loss of FMRP, an mRNA-binding protein that negatively regulates translation^{37,38}. In the *Fmr1*^{-/-} mouse model, basal protein synthesis is elevated and LTD is exaggerated downstream of an mGluR5 signalling pathway involving ERK1/2 (ref. 23). Partial inhibition of mGluR5 corrects multiple aspects of fragile X in animal models^{2,39}. Recent data suggest that the mTOR signalling pathway is also constitutively overactive in the *Fmr1*^{-/-} mouse¹⁷, but the relevance to exaggerated protein synthesis and altered synaptic function has been unclear. The current findings show that increased synaptic mTOR activity actually suppresses the protein synthesis required for LTD in the *Tsc2*^{+/-} mice. The idea that reduced protein synthesis is a causative factor in the observed deficit in synaptic plasticity is supported by the finding that pharmacological rescue with both rapamycin and CDPPB is abolished by cycloheximide, and the observation that Arc is deficiently translated in the *Tsc2*^{+/-} mice. There is good evidence that Arc is one of the proteins that normally must be synthesized to support mGluR5-dependent forms of long-term plasticity^{29,30}. Precisely how excess mTOR activity suppresses synthesis of these plasticity proteins remains to be investigated, but possibilities include hyperphosphorylation of FMRP⁴⁰ or increased translation of a competing pool of less abundant mRNAs unrelated to LTD^{7,21,30}. The fact that mGluR-LTD is altered in opposite directions in *Tsc2*^{+/-} and *Fmr1*^{-/-} mice, and that both deviations are corrected in the double mutants, suggests that the pool comprising LTD proteins is differentially regulated by FMRP and TSC1/2 (Fig. 3a).

The current findings also suggest a new treatment for behavioural deficits associated with TSC. Previous studies in the *Tsc2*^{+/-} mouse raised the exciting possibility that cognitive aspects of the disorder might be ameliorated with rapamycin, even when treatment is begun in adulthood⁴. Our data show that an mGluR5 PAM may be similarly effective. Although rapamycin has been used clinically, it is problematic for chronic treatment because of its strong immunosuppressive properties. The benefit of mGluR5 PAMs is that they target specifically the synaptic mechanisms that are probably responsible for the cognitive and behavioural impairments in TSC.

TSC and FXS represent two leading genetic risk factors for ASD and intellectual disability⁴¹. Although great strides have been made in identifying genetic variation that correlates with non-syndromic autism, little is known about ASD pathophysiology, knowledge that is essential for developing effective therapies. Our test of the hypothesis that the *Fmr1*^{-/-} and *Tsc2*^{+/-} models of FXS and TSC have a shared synaptic pathophysiology revealed instead that they are at opposite ends of a spectrum: the *Fmr1* mutation causes exaggerated synaptic protein synthesis and LTD that are corrected by inhibition of mGluR5 (ref. 31), whereas the *Tsc2* mutation causes diminished synaptic protein synthesis and LTD that are corrected by augmentation of mGluR5 (Fig. 4a). Moreover, the opposing effects of these mutations balance one another at synaptic and behavioural levels in the double mutant. This finding is interesting in light of recent discoveries that gain- and loss-of-function mutations in individual genes, such as *MECP2*, can often yield syndromes with overlapping features, such as epilepsy, cognitive impairment and ASD⁴². Our findings reveal that even genetically heterogeneous causes of ASD and intellectual disability may produce similar deficits by bidirectional deviations from normal on a common functional axis. The important implication is that therapies designed to correct one cause of ASD are not likely to be effective for all other causes, and might well be deleterious. It will be critical to understand where a patient lies on the spectrum of synaptic function to choose an appropriate therapy for ASD and other psychiatric disorders.

METHODS SUMMARY

All experimental procedures were approved by the Institutional Animal Care and Use Committee at Massachusetts Institute of Technology. Age-matched, male littermate mice bred on the C57Bl/6J clonal background were used in this study.

Acute hippocampal slices were prepared from P25–35 mice as previously described³¹ and field excitatory postsynaptic potentials (fEPSPs) evoked by stimulation of the Schaffer collaterals were recorded in CA1 stratum radiatum with extracellular electrodes. LTD was induced by applying R,S-DHPG (50 μ M) or S-DHPG (25 μ M) for 5 min, or by paired-pulse low-frequency synaptic stimulation^{18,19} for 20 min. Metabolic labelling, immunoblotting and immunoprecipitation experiments were performed on yoked WT and *Tsc2*^{+/-} mice as described previously^{23,43}. Context discrimination fear conditioning was performed as described previously⁴. For *in vivo* mGluR5 PAM experiments, animals received a single injection of CDPBB (10 mg kg⁻¹, intraperitoneal) 30 min before the training session. For all data sets, outliers more than 2 σ from the mean were removed, and significance between more than two groups was determined using two-way analysis of variance (ANOVA) and *post hoc* Student's *t*-tests. Statistics were performed using each animal as an 'n', with each animal represented by the mean of one to four slices for electrophysiology experiments, one slice per animal for biochemistry experiments and five to eight slices per animal for immunoprecipitation experiments. All experiments were performed blind to genotype and include interleaved controls for genotype and treatment. For detailed methods, see Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 24 August; accepted 21 October 2011.

Published online 23 November 2011.

- Newschaffer, C. J. *et al.* The epidemiology of autism spectrum disorders. *Annu. Rev. Public Health* **28**, 235–258 (2007).
- Krueger, D. D. & Bear, M. F. Toward fulfilling the promise of molecular medicine in fragile X syndrome. *Annu. Rev. Med.* **62**, 411–429 (2011).
- Kelleher, R. J. III & Bear, M. F. The autistic neuron: troubled translation? *Cell* **135**, 401–406 (2008).
- Ehninger, D. *et al.* Reversal of learning deficits in a *Tsc2*^{+/-} mouse model of tuberous sclerosis. *Nature Med.* **14**, 843–848 (2008).
- Meikle, L. *et al.* Response of a neuronal model of tuberous sclerosis to mammalian target of rapamycin (mTOR) inhibitors: effects on mTORC1 and Akt signaling lead to improved survival and function. *J. Neurosci.* **28**, 5422–5432 (2008).
- Onda, H. *et al.* *Tsc2* null murine neuroepithelial cells are a model for human tuber giant cells, and show activation of an mTOR pathway. *Mol. Cell. Neurosci.* **21**, 561–574 (2002).
- Ehninger, D., de Vries, P. J. & Silva, A. J. From mTOR to cognition: molecular and cellular mechanisms of cognitive impairments in tuberous sclerosis. *J. Intellect. Disabil. Res.* **53**, 838–851 (2009).
- Carbonara, C. *et al.* 9q34 loss of heterozygosity in a tuberous sclerosis astrocytoma suggests a growth suppressor-like activity also for the *TSC1* gene. *Hum. Mol. Genet.* **3**, 1829–1832 (1994).
- Green, A. J., Smith, M. & Yates, J. R. Loss of heterozygosity on chromosome 16p13.3 in hamartomas from tuberous sclerosis patients. *Nature Genet.* **6**, 193–196 (1994).
- de Vries, P. J. & Howe, C. J. The tuberous sclerosis complex proteins—a GRIPP on cognition and neurodevelopment. *Trends Mol. Med.* **13**, 319–326 (2007).
- Goorden, S. M., van Woerden, G. M., van der Weerd, L., Cheadle, J. P. & Elgersma, Y. Cognitive deficits in *Tsc1*^{+/-} mice in the absence of cerebral lesions and seizures. *Ann. Neurol.* **62**, 648–655 (2007).
- Cheadle, J. P., Reeve, M. P., Sampson, J. R. & Kwiatkowski, D. J. Molecular genetic advances in tuberous sclerosis. *Hum. Genet.* **107**, 97–114 (2000).
- Onda, H., Lueck, A., Marks, P. W., Warren, H. B. & Kwiatkowski, D. J. *Tsc2*^{+/-} mice develop tumors in multiple sites that express gelsolin and are influenced by genetic background. *J. Clin. Invest.* **104**, 687–695 (1999).
- Nie, D. *et al.* *Tsc2*-Rheb signaling regulates EphA-mediated axon guidance. *Nature Neurosci.* **13**, 163–172 (2010).
- Young, D. M., Schenk, A. K., Yang, S. B., Jan, Y. N. & Jan, L. Y. Altered ultrasonic vocalizations in a tuberous sclerosis mouse model of autism. *Proc. Natl Acad. Sci. USA* **107**, 11074–11079 (2010).
- Hoeffer, C. A. & Klann, E. mTOR signaling: at the crossroads of plasticity, memory and disease. *Trends Neurosci.* **33**, 67–75 (2010).
- Sharma, A. *et al.* Dysregulation of mTOR signaling in fragile X syndrome. *J. Neurosci.* **30**, 694–702 (2010).
- Huber, K. M., Kayser, M. S. & Bear, M. F. Role for rapid dendritic protein synthesis in hippocampal mGluR-dependent long-term depression. *Science* **288**, 1254–1257 (2000).
- Huber, K. M., Roder, J. C. & Bear, M. F. Chemical induction of mGluR5- and protein synthesis-dependent long-term depression in hippocampal area CA1. *J. Neurophysiol.* **86**, 321–325 (2001).
- Huber, K. M., Gallagher, S. M., Warren, S. T. & Bear, M. F. Altered synaptic plasticity in a mouse model of fragile X mental retardation. *Proc. Natl Acad. Sci. USA* **99**, 7746–7750 (2002).
- Bear, M. F., Huber, K. M. & Warren, S. T. The mGluR theory of fragile X mental retardation. *Trends Neurosci.* **27**, 370–377 (2004).
- Gallagher, S. M., Daly, C. A., Bear, M. F. & Huber, K. M. Extracellular signal-regulated protein kinase activation is required for metabotropic glutamate receptor-dependent long-term depression in hippocampal area CA1. *J. Neurosci.* **24**, 4859–4864 (2004).
- Osterweil, E. K., Krueger, D. D., Reinhold, K. & Bear, M. F. Hypersensitivity to mGluR5 and ERK1/2 leads to excessive protein synthesis in the hippocampus of a mouse model of fragile X syndrome. *J. Neurosci.* **30**, 15616–15627 (2010).
- Fitzjohn, S. M. *et al.* A characterisation of long-term depression induced by metabotropic glutamate receptor activation in the rat hippocampus *in vitro*. *J. Physiol.* **537**, 421–430 (2001).
- Nosyreva, E. D. & Huber, K. M. Developmental switch in synaptic mechanisms of hippocampal metabotropic glutamate receptor-dependent long-term depression. *J. Neurosci.* **25**, 2992–3001 (2005).
- Mockett, B. G. *et al.* Calcium/calmodulin-dependent protein kinase II mediates group I metabotropic glutamate receptor-dependent protein synthesis and long-term depression in rat hippocampus. *J. Neurosci.* **31**, 7380–7391 (2011).
- Luscher, C. & Huber, K. M. Group 1 mGluR-dependent synaptic long-term depression: mechanisms and implications for circuitry and disease. *Neuron* **65**, 445–459 (2010).
- Snyder, E. M. *et al.* Internalization of ionotropic glutamate receptors in response to mGluR activation. *Nature Neurosci.* **4**, 1079–1085 (2001).
- Wang, M. W., Pfeiffer, B. E., Nosyreva, E. D., Ronesi, J. A. & Huber, K. M. Rapid translation of Arc/Arg3.1 selectively mediates mGluR-dependent LTD through persistent increases in AMPAR endocytosis rate. *Neuron* **59**, 84–97 (2008).
- Park, S. *et al.* Elongation factor 2 and fragile X mental retardation protein control the dynamic translation of Arc/Arg3.1 essential for mGluR-LTD. *Neuron* **59**, 70–83 (2008).
- Dolen, G. *et al.* Correction of fragile X syndrome in mice. *Neuron* **56**, 955–962 (2007).
- Conn, P. J., Christopoulos, A. & Lindsley, C. W. Allosteric modulators of GPCRs: a novel approach for the treatment of CNS disorders. *Nature Rev. Drug Discov.* **8**, 41–54 (2009).
- Kinney, G. G. *et al.* A novel selective positive allosteric modulator of metabotropic glutamate receptor subtype 5 has *in vivo* activity and antipsychotic-like effects in rat behavioral models. *J. Pharmacol. Exp. Ther.* **313**, 199–206 (2005).
- Frankland, P. W., Cestari, V., Filipkowski, R. K., McDonald, R. J. & Silva, A. J. The dorsal hippocampus is essential for context discrimination but not for contextual conditioning. *Behav. Neurosci.* **112**, 863–874 (1998).
- Lu, Y. M. *et al.* Mice lacking metabotropic glutamate receptor 5 show impaired learning and reduced CA1 long-term potentiation (LTP) but normal CA3 LTP. *J. Neurosci.* **17**, 5196–5205 (1997).
- Stiedl, O., Palve, M., Radulovic, J., Birkenfeld, K. & Spiess, J. Differential impairment of auditory and contextual fear conditioning by protein synthesis inhibition in C57BL/6N mice. *Behav. Neurosci.* **113**, 496–506 (1999).
- Bassell, G. J. & Warren, S. T. Fragile X syndrome: loss of local mRNA regulation alters synaptic development and function. *Neuron* **60**, 201–214 (2008).
- Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146**, 247–261 (2011).
- Dolen, G., Carpenter, R. L., Ocain, T. D. & Bear, M. F. Mechanism-based approaches to treating fragile X. *Pharmacol. Ther.* **127**, 78–93 (2010).
- Narayanan, U. *et al.* S6K1 phosphorylates and regulates fragile X mental retardation protein (FMRP) with the neuronal protein synthesis-dependent mammalian target of rapamycin (mTOR) signaling cascade. *J. Biol. Chem.* **283**, 18478–18482 (2008).
- Fombonne, E. Epidemiological surveys of autism and other pervasive developmental disorders: an update. *J. Autism Dev. Disord.* **33**, 365–382 (2003).
- Ramocki, M. B. & Zoghbi, H. Y. Failure of neuronal homeostasis results in common neuropsychiatric phenotypes. *Nature* **455**, 912–918 (2008).
- Krueger, D. D., Osterweil, E. K. & Bear, M. F. Activation of mGluR5 induces rapid and long-lasting protein kinase D phosphorylation in hippocampal neurons. *J. Mol. Neurosci.* **42**, 1–8 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was partly supported by grants from the National Institute of Mental Health (T32 MH-082718 and T32-MH-074249), the National Institute of Child Health and Human Development (2R01HD046943), the Department of Defense (W81XWH-11-1-0252) and The Simons Foundation. We acknowledge A. Heynen for advice and comments, as well as K. Oram, E. Sklar and S. Meagher for technical and administrative assistance. Monoclonal Arc antibody was a gift from P. Worley.

Author Contributions M.F.B. designed, directed and coordinated the project. B.D.A. designed and performed electrophysiological recordings and behaviour tests. E.K.O. designed and performed biochemistry experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.F.B. (mbear@mit.edu).

METHODS

Animals. $Tsc2^{+/-}$ male and female mutant mice on the C57Bl/6J clonal background were bred with C57Bl/6J WT partners to produce the WT and $Tsc2^{+/-}$ male offspring used in this study. For genetic rescue experiments, heterozygous $Tsc2$ male mice ($Tsc2^{+/-}$) were bred with heterozygous $Fmr1$ females ($Fmr1^{x^+/x^-}$), both on the C57Bl/6J clonal background, to obtain F1 male offspring of four genotypes: WT ($Tsc2^{+/+}$, $Fmr1^{+/+}$), $Fmr1$ KO ($Tsc2^{+/+}$, $Fmr1^{-/-}$), $Tsc2$ Het ($Tsc2^{+/-}$, $Fmr1^{+/+}$) and cross ($Tsc2^{+/-}$, $Fmr1^{-/-}$) (Fig. 4b). All experimental animals were age-matched male littermates, and were studied with the experimenter blind to genotype and treatment condition. Animals were group housed and maintained on a 12:12 h light:dark cycle. The Institutional Animal Care and Use Committee at Massachusetts Institute of Technology approved all experimental techniques.

Electrophysiology. Acute hippocampal slices were prepared from P25–35 animals in ice-cold dissection buffer containing (in mM): NaCl 87, sucrose 75, KCl 2.5, NaH_2PO_4 1.25, $NaHCO_3$ 25, $CaCl_2$ 0.5, $MgSO_4$ 7, ascorbic acid 1.3, and D-glucose 10 (saturated with 95% O_2 /5% CO_2). Immediately after slicing, the CA3 region was removed. Slices were recovered in artificial cerebrospinal fluid (ACSF) containing (in mM): NaCl 124, KCl 5, NaH_2PO_4 1.23, $NaHCO_3$ 26, $CaCl_2$ 2, $MgCl_2$ 1 and D-glucose 10 (saturated with 95% O_2 /5% CO_2) at 32.5 °C for at least 3 h before recording.

Field recordings were performed in a submersion chamber, perfused with ACSF (2–3 ml min⁻¹) at 30 °C. fEPSPs were recorded in CA1 stratum radiatum with extracellular electrodes filled with ACSF. Baseline responses were evoked by stimulation of the Schaffer collaterals at 0.033 Hz with a two-contact cluster electrode (FHC) using a 0.2 ms stimulus yielding 40–60% of the maximal response. Field potential recordings were filtered at 0.1 Hz to 1 kHz, digitized at 10 kHz and analysed using pClamp9 (Axon Instruments). The initial slope of the response was used to assess changes in synaptic strength. Data were normalized to the baseline response and are presented as group means \pm s.e.m. LTD was measured by comparing the average response 55–60 min after DHPG application to the average of the last 5 min of baseline.

The input–output function was examined by stimulating slices with incrementally increasing current and recording the fEPSP response. Paired-pulse facilitation was induced by applying two pulses at different inter-stimulus intervals. Facilitation was measured by the ratio of the fEPSP slope of stimulus 2 to stimulus 1. NMDAR-dependent LTD was induced by delivering 900 test pulses at 1 Hz. mGluR-LTD was induced by applying R,S-DHPG (50 μ M) or S-DHPG (25 μ M) for 5 min, or by delivering 1,200 pairs of pulses (with a 50 ms interstimulus interval) at 1 Hz. In some experiments, slices were incubated with the protein synthesis inhibitor cycloheximide (60 μ M) for 30 min as follows: 20 min during baseline recording, 5 min during DHPG application and 5 min after DHPG application. For mGluR5 PAM experiments, slices were pretreated with CDPPB (10 μ M) or DMSO control for 30 min in same manner as above, either in the presence of cycloheximide or control ACSF. For rapamycin experiments, slices were pretreated with rapamycin (20 nM) or DMSO control, with or without cycloheximide, for at least 30 min before recording and throughout the entire experiment. Significance was determined by two-way ANOVA and *post hoc* Student's *t*-tests. Statistics were performed using each animal as an 'n', with each animal represented by the mean of one to four slices. All experiments were performed blind to genotype and included interleaved controls for genotype and treatment.

Metabolic labelling of new protein synthesis. Performed as described by Osterweil *et al.*²³. Briefly, 500 μ m slices were recovered for 4 h in 32.5 °C ACSF (in mM: 124 NaCl, 3 KCl, 1.25 NaH_2PO_4 , 26 $NaHCO_3$, 10 dextrose, 1 $MgCl_2$, 2 $CaCl_2$, saturated with 95% O_2 and 5% CO_2), incubated for 30 min with 25 μ M ActD \pm rapamycin (20 nM) or CDPPB (10 μ M), and transferred to fresh ACSF \pm drug with 10 μ Ci ml⁻¹ [³⁵S]Met/Cys (Perkin Elmer) for another

30 min. After labelling, slices were homogenized, and labelled proteins isolated by TCA precipitation. Samples were read with a scintillation counter and subjected to a protein concentration assay (Bio-Rad). Final data were expressed as counts per minute per microgram of protein, normalized to the [³⁵S]Met/Cys ACSF used for incubation and the average incorporation of all samples analysed in that experiment. For autoradiography, homogenized slices were processed for SDS–polyacrylamide gel electrophoresis, transferred to nitrocellulose, stained for total protein using a Memcode staining kit (Pierce) and ³⁵S-incorporated proteins visualized with the aid of a phosphorimager (Fujifilm).

Immunoblotting. Immunoblotting was performed according to established methods using primary antibodies to Arc (Synaptic Systems), p-ERK1/2 (Thr 202/Tyr 204) (Cell Signaling Technology) or ERK1/2 (Cell Signaling Technology). ERK1/2 phosphorylation was measured by densitometry (Quantity One), and quantified as the densitometric signal of p-ERK1/2 divided by the ERK1/2 signal in the same lane. To quantify Arc expression, the densitometric signal of Arc was divided by the total protein signal (determined by Memcode staining) in the same lane.

Immunoprecipitation. Hippocampal slices (five to eight per animal) were metabolically labelled with 50 μ Ci ml⁻¹ [³⁵S]Met/Cys for 3 h, and immunoprecipitation performed on yoked WT and $Tsc2^{+/-}$ slices essentially as described previously²³. Briefly, slices were homogenized in immunoprecipitation lysis buffer (Pierce) plus protease inhibitors (EMD Biosciences), spun at 16,000g and supernatants pre-cleared with protein A/G sepharose. To avoid contamination of the Arc signal with IgG heavy chain, immunoprecipitation was performed using columns of monoclonal Arc antibody (a gift from P. Worley) crosslinked to protein A/G sepharose (Pierce Crosslink IP Kit). Immunoprecipitated Arc was resolved on SDS–polyacrylamide gel electrophoresis gels, transferred to nitrocellulose and exposed to a phosphorimager screen for 2–3 weeks. The same membranes were then immunoblotted for Arc. For each sample, the ratio of ³⁵S-incorporated: total was calculated by dividing the density of the band seen by autoradiography to the density of band seen by immunoblot (in the same lane).

Contextual fear conditioning. Six- to 12-week-old WT, $Tsc2^{+/-}$, $Fmr1^{-/-}$ and cross ($Tsc2^{+/-} \times Fmr1^{-/-}$) mice were fear conditioned to the training context with one 0.8 mA shock (2 s) as described by Ehninger *et al.*⁴. The mice were allowed 3 min to explore context before conditioning and were removed 15 s after the shock was given and returned to the home cage. Conditioned fear response was assessed 24 h later by a trained observer blind to condition, measuring the percentage of time spent freezing during the test period (3 min session). To determine context specificity of the conditioned response, mice trained at the same time were separated into two groups: one group was tested in the same training context, the other tested in a novel context. This novel context was created by varying distal cues, odour (2% acetic acid compared with 70% ethanol), floor material (plastic compared with metal bars) and lighting (red compared with white) of the testing apparatus. For rescue experiments, animals received a single injection of CDPPB (10 mg kg⁻¹, intraperitoneal) 30 min before training session.

Reagents: R,S-DHPG was purchased from Tocris Biosciences and S-DHPG was purchased from Sigma. Fresh bottles of DHPG were prepared as a 100 \times stock in H₂O, divided into aliquots and stored at –80 °C. Fresh stocks were made once a week. Rapamycin (EMD Biosciences) was prepared at 10 mM stock in DMSO and stored at –80 °C. The final concentration of rapamycin was 20 nM in less than 0.01% DMSO. Cycloheximide (Sigma) was prepared daily at 100 \times stock in H₂O. For slice experiments, CDPPB (EMD Biosciences) was prepared daily at 75 mM stock in DMSO with 0.5% bovine serum albumin (BSA) and diluted in ACSF to achieve a final concentration of 10 μ M in less than 0.1% DMSO. For *in vivo* experiments, CDPPB was suspended in a vehicle consisting of 20% (2-hydroxypropyl)-(R)-cyclodextrin in sterile saline. All other reagents were purchased from Sigma.

The unusual gamma-ray burst GRB 101225A explained as a minor body falling onto a neutron star

S. Campana¹, G. Lodato², P. D'Avanzo¹, N. Panagia^{3,4,5}, E. M. Rossi⁶, M. Della Valle⁷, G. Tagliaferri¹, L. A. Antonelli⁸, S. Covino¹, G. Ghirlanda¹, G. Ghisellini¹, A. Melandri¹, E. Pian^{9,10}, R. Salvaterra¹¹, G. Cusumano¹², V. D'Elia^{13,8}, D. Fugazza¹, E. Palazzi¹⁴, B. Sbarufatti¹ & S. D. Vergani¹

The tidal disruption of a solar-mass star around a supermassive black hole has been extensively studied analytically^{1,2} and numerically³. In these events, the star develops into an elongated banana-shaped structure. After completing an eccentric orbit, the bound debris falls into the black hole, forming an accretion disk and emitting radiation^{4–6}. The same process may occur on planetary scales if a minor body passes too close to its star. In the Solar System, comets fall directly into our Sun⁷ or onto planets⁸. If the star is a compact object, the minor body can become tidally disrupted. Indeed, one of the first mechanisms invoked to produce strong gamma-ray emission involved accretion of comets onto neutron stars in our Galaxy⁹. Here we report that the peculiarities of the ‘Christmas’ gamma-ray burst (GRB 101225A¹⁰) can be explained by a tidal disruption event of a minor body around an isolated Galactic neutron star. This would indicate either that minor bodies can be captured by compact stellar remnants more frequently than occurs in the Solar System or that minor-body formation is relatively easy around millisecond radio pulsars. A peculiar supernova associated with a gamma-ray burst provides an alternative explanation¹¹.

GRB 101225A image-triggered the Burst Alert Telescope (BAT) on board NASA's Swift satellite on 25.776 December 2010 UT. The event was extremely long, with a T_{90} (the time interval during which 90% of the flux was emitted) of >1.7 ks, and smooth¹⁰ by comparison with typical gamma-ray bursts¹² (GRBs). The total 15–150-keV fluence recorded by the BAT is $\gtrsim 3 \times 10^{-6}$ erg cm⁻² and there are no signs of decay. The X-ray Telescope and the Ultraviolet–Optical Telescope on board Swift found a bright, long-lasting counterpart to the GRB. Strong variability is observed in the early X-ray light curve. The optical counterpart, which was detected in all of the Ultraviolet–Optical Telescope filters, lags the X-ray light curve (Fig. 1). The X-ray and optical light curves are reminiscent of the shock break-out event observed in association with GRB 060218¹³, but are fainter (~ 3.5 mag; that is, fainter by a factor of ~ 25) and do not have an X-ray afterglow at later times or a bright supernova component. Measurements by the European Space Agency's XMM-Newton space observatory failed to detect an afterglow with an upper flux limit of $\sim 10^{-14}$ erg cm⁻² s⁻¹ at the 3σ confidence level (0.5–10 keV; observation made at $\Delta t = 23$ d after the trigger).

Ground-based telescopes also followed the event, mainly in the R and I bands (Fig. 1). Surprisingly, optical spectra recorded in the first few days after the event failed to detect any spectral features, instead revealing a smooth blue continuum^{11,14}. At later times, the optical light curves revealed a colour change from blue to red. The NASA Hubble Space Telescope imaged the field at $\Delta t = 20$ d, finding a quite red object ($V_{606}(\text{AB}) = 24.6$ mag and $B_{435} - V_{606} = 1.6$ mag), with no host

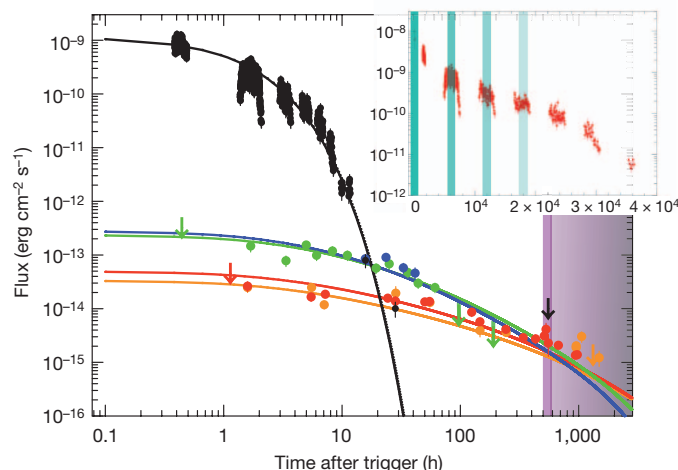


Figure 1 | Light curves of GRB 101225A. GRB 101225A light curves in five energy bands: X-rays at 1 keV (black), ultraviolet at 2,030 Å (green) and 2,634 Å (blue), and optical at 6,400 Å (R band, red) and 7,700 Å (I band, orange). Error bars, 1σ . Black arrow indicates the XMM-Newton upper limit. Other arrows indicate UV and optical upper limits according to their colour coding. The X-ray light curve represents only the disk contribution to the total flux (~ 0.3 of the total, as derived from spectral modelling) and is corrected for the interstellar absorption column density, of $N_{\text{H}} = (2.0 \pm 0.1) \times 10^{21}$ cm⁻² (which is greater than the Galactic value, $N_{\text{H}}^{\text{Gal}} = 7.9 \times 10^{20}$ cm⁻²). This enhanced column density may be due to the fraction of the minor body's mass that has been expelled from the system. The continuous lines of different colours (the same as the data) represent the fit to the light curves using the phenomenological model of tidal disruption²¹. Because the model predicts a late transition of the accretion disk to a ‘cold’ solution, the fit has been carried out up to ~ 20 d (the excluded region is indicated in pink-grey). A thick pink line indicates the time of the transition. Our model has four parameters: the mass of the minor body (M_* ; we assume for simplicity a density of 1 g cm^{-3} , thereby fixing the radius, R_*), the periastron (r_p), a geometrical factor ($D^2/\cos(i)$, where D is the source distance and i the source inclination) and the optical absorption (A_V). The best fit is obtained for $M_* \approx 5 \times 10^{20} \text{ g}$, $r_p \approx 9 \times 10^3 \text{ km}$, $D/\sqrt{\cos(i)} \approx 3 \text{ kpc}$ and $A_V \approx 0.8$ (in excess of the Galactic value, $A_V^{\text{Gal}} = 0.3$, consistent with the value determined by X-ray analysis). The peak mass accretion rate with these parameters is $\dot{M} \approx 2 \times 10^{16} \text{ g s}^{-1}$ and the peak luminosity is $L \approx 3 \times 10^{36} \text{ erg s}^{-1}$, consistent with our hypothesis of sub-Eddington accretion. In this regime, no emission lines are expected, at variance with super-Eddington events. Inset, X-ray light curve as observed by the Swift X-ray Telescope. Shaded regions highlight the periastron passages calculated using our tidal disruption model, which have a timescale of 6,000–10,000 s. Outside each band, the X-ray emission markedly decreases, and this effect is strongest for the first passage. We verified that this effect is not instrumental in nature nor due to spectral changes. Inhomogeneities in the return debris are expected to give rise to variability on the fall-back timescale.

¹INAF - Osservatorio Astronomico di Brera, Via E. Bianchi 46, I-23807 Merate (LC), Italy. ²Dipartimento di Fisica, Università degli Studi di Milano, Via Celoria 16, I-20133 Milano, Italy. ³Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. ⁴INAF - Osservatorio Astrofisico di Catania, Via Santa Sofia 78, I-95123 Catania, Italy. ⁵Supernova Ltd, OYV #131, Northsound Road, Virgin Gorda, British Virgin Islands. ⁶Leiden Observatory, Leiden University, PO Box 9513, 2300 RA Leiden, The Netherlands. ⁷INAF - Osservatorio Astronomico di Capodimonte, Salita Moiarriello 16, I-80131 Napoli, Italy. ⁸INAF - Osservatorio Astronomico di Roma, Via Frascati 33, I-00044 Monte Porzio (Roma), Italy. ⁹Scuola Normale Superiore, Piazza dei Cavalieri 7, I-56126 Pisa, Italy. ¹⁰INAF - Osservatorio Astronomico di Trieste, Via G. Tiepolo 11, I-34143 Trieste, Italy. ¹¹Dipartimento di Fisica e Matematica, Università dell'Insubria, Via Valleggio 7, I-22100 Como, Italy. ¹²INAF - Istituto di Astrofisica Spaziale e Fisica Cosmica, Via U. La Malfa 153, I-90146 Palermo, Italy. ¹³ASI Science Data Centre, Via Galileo Galilei, I-00044 Frascati (Roma), Italy. ¹⁴INAF - Istituto di Astrofisica Spaziale e Fisica Cosmica, Sezione di Bologna, Via Gobetti 101, I-40129 Bologna, Italy.

galaxy apparent at magnitudes at most 1 mag fainter¹⁵. Later observations carried out at the INAF Telescopio Nazionale Galileo recorded an I-band magnitude of $I = 23.64 \pm 0.24$ mag at $\Delta t = 44$ d but recorded a non-detection ($I > 24.5$ mag, 3σ confidence level) at $\Delta t = 55$ d. Observations at the Spanish Gran Telescopio Canarias ($\Delta t = 180$ d) detected GRB 101225A at $g'_{AB} = 27.21 \pm 0.27$ mag and $r'_{AB} = 26.90 \pm 0.14$ mag (ref. 11).

These observational characteristics make GRB 101225A unique. Without a-priori knowledge of the distance, this event can result from different progenitors. We explored several possibilities (Supplementary Information). Extragalactic models (involving GRBs, supernovae or blazars) are unlikely on the basis of total energy arguments, in combination with the faintness of the optical counterpart and putative host galaxy. Galactic models (involving X-ray binaries, magnetars or flare stars) are ruled out as well, on the basis of the extremely faint quiescent counterpart and the relatively high Galactic latitude. The timescale of the event, the longer duration at longer wavelengths and the faintness of a quiescent counterpart at all wavelengths all suggest a disruptive event involving a compact object. We propose that GRB 101225A results from the tidal disruption of a minor body around an isolated neutron star. Similar models have been already put forward to explain the presence of metal absorption lines in the white dwarf optical spectra¹⁶ and the infrared excess in the spectra of hot white dwarfs¹⁷.

A minor body gets disrupted if it comes within a distance of $\sim 10^5$ – 10^6 km from the neutron star, where the internal forces that hold the body together are overwhelmed by the tidal pull of the neutron star. The debris that remain bound are thrown into highly eccentric orbits and then fall back to form an accretion disk around the star^{18,19}. If the minor body has a low tensile strength, it gets fully disrupted (breaking down into gas), its early behaviour can be chaotic owing to the formation of the disk with flares coincident with the periastron passages^{6,20}, and the bolometric light curve is expected to decay, in the long term, as $t^{-5/3}$ (refs 1, 2).

If the matter that falls back to form the disk accretes at a sub-Eddington rate at all times (as in our case), the disk emission can be described as a spatial superposition of black bodies of increasing temperatures and depends on the parameters of the encounter²¹. The physics of the event depends, to a first approximation, only on a very few parameters: the periastron of the minor body's orbit (r_p), the minor body's mass and radius (M_* and R_* ; the mean density of a minor body in the Solar System is in the range ~ 1 – 10 g cm⁻³, which adds an additional constraint), the minor body's tensile strength, the neutron star mass (M_{NS}) and the source distance (the distance between the neutron star and Earth).

We fit the X-ray (1 keV), ultraviolet (UVW2 and UVW1 bands, centred at 2,030 and 2,634 Å, respectively) and optical (R and I bands, centred at 6,400 and 7,700 Å, respectively) light curves of GRB 101225A with a four-parameter phenomenological model of tidal disruption onto a neutron star²¹ (Fig. 1). We assume that the neutron star mass is 1.4 solar masses and that the minor body has a density of 1 g cm⁻³ and a low internal tensile strength. Our model is indicative of the physics involved, even though it cannot capture the full details of the event. We find an adequate fit for $M_* \approx 5 \times 10^{20}$ g (half the mass of the asteroid Ceres) and a periastron radius of $\sim 9,000$ km (which is well within the tidal radius, thus demonstrating the internal consistency of the model). The mass of the minor body can be estimated within a factor of $\lesssim 3$. This mass is large by comparison with those of objects in the asteroid belt but is typical of objects in the Kuiper belt.

Assuming that half of the mass is accreted onto the neutron star and that the total fluence is 4×10^{-5} erg cm⁻², estimated within the 1 – 10^5 -s time interval (including a bolometric correction motivated by spectral fitting), we derive a distance of ~ 3 kpc, placing GRB 101225A in the Perseus arm of the Galaxy. The observed radiation is emitted both by an accretion disk (with a outer radius twice as large as periastron) and during the impact of matter on the neutron star surface. The early

spectral energy distribution (SED) at five different epochs can be adequately fitted by the emission from an accretion disk (with a temperature of $T_d \approx 1.8$ keV in units of energy, in agreement with model predictions) plus the emission from a boundary layer in which the disk matter slows down to accrete onto the neutron star (Supplementary Information). In particular, the initial high-energy emission observed by the BAT can be modelled as a black body with $T \approx 10$ keV (ref. 11). This cannot come from the disk but can be easily accounted for by mass accretion onto the neutron star surface. The presence of the boundary layer component rules out the compact object's being a black hole. Small differences in the return times of debris are expected during the first periastron passages, and these give rise to X-ray variability on the timescale of this fall back^{6,20,22}, as observed in the X-ray band (Fig. 1, inset). The X-ray variability timescale is indeed very similar to the fall-back timescale obtained from our modelling.

The phenomenological model can explain the (observed) slower decay at longer wavelengths. It somewhat underestimates the observed decay of the optical flux starting at ~ 20 d. Changes in the properties of the accretion disk are expected on the basis of the theory of accretion onto neutron stars. The inner disk radius is increased by the neutron star's magnetic field (even when this takes its minimal value, 10^8 – 10^9 G). Furthermore, when the accretion rate decreases below a critical value, the disk develops a well-known thermal–viscous instability leading to fast cooling^{23,24}. We have calculated that, on the basis of the predicted evolution of the mass inflow rate, this transition occurs at $\Delta t \approx 15$ – 21 d. The $\Delta t = 40$ d SED can be modelled as arising from a cool ring ($T \approx 4,500$ K) with a large inner radius. The $\Delta t = 180$ d SED can be interpreted as the emission coming from a larger and colder ring ($T \approx 650$ K, resembling a protoplanetary disk) plus a contribution from the neutron star heated by the accreted material (Supplementary Information).

Estimating the rate of such events is challenging. A wandering neutron star can intercept minor bodies as it passes through a planetary system. The capture rate can be evaluated on the basis of the hypothesis that all stars have Oort clouds similar to that of the Sun²⁵. Applying this to the case of GRB 101225A, we derive a rate of ~ 0.3 events per year, in line with Swift observations. However, indirect arguments suggest a lower capture rate²⁶ and, in addition, larger minor bodies are rarer. The likelihood of a neutron star retaining its original population of small bodies is low, because they are unlikely to survive the supernova event. Nevertheless, planetary-like systems can reform around millisecond radio pulsars^{27,28}, providing a viable mechanism for the occurrence of GRB 101225A.

Received 8 April; accepted 23 September 2011.

1. Rees, M. J. Tidal disruption of stars by black holes of 10^6 – 10^8 solar masses in nearby galaxies. *Nature* **333**, 523–528 (1988).
2. Phinney, E. S. in *Proc. 136th IAU Symp. 'The Center of the Galaxy'* (ed. Morris, M.) 543–553 (Kluwer, 1989).
3. Evans, C. R. & Kochanek, C. S. The tidal disruption of a star by a massive black hole. *Astrophys. J.* **346**, L13–L16 (1989).
4. Renzini, A. et al. An ultraviolet flare at the centre of the elliptical galaxy NGC4552. *Nature* **378**, 39–41 (1995).
5. Bade, N., Komossa, S. & Dahlem, M. Detection of an extremely soft X-ray outburst in the HII-like nucleus of NGC 5905. *Astron. Astrophys.* **309**, L35–L38 (1996).
6. Bloom, J. S. et al. A possible relativistic jetted outburst from a massive black hole fed by a tidally disrupted star. *Science* **333**, 203–206 (2011).
7. Sheeley, N. R. Jr, Howard, R. A., Koomen, M. J. & Michels, D. J. Coronagraphic observations of two new sungrazing comets. *Nature* **300**, 239–242 (1982).
8. Harrington, J., LeBeau, R. P. Jr, Backes, K. A. & Dowling, T. E. Dynamic response of Jupiter's atmosphere to the impact of comet Shoemaker-Levy 9. *Nature* **368**, 525–527 (1994).
9. Harwit, M. & Salpeter, E. E. Radiation from comets near neutron stars. *Astrophys. J.* **186**, L37–L39 (1973).
10. Racusin, J. L. et al. Swift observation of GRB101225A. *GCN Rep.* **314.1** (2011).
11. Thöne, C. C. et al. The unusual γ -ray burst GRB 101225A from a helium star/neutron star merger at redshift 0.33. *Nature* doi:10.1038/nature10611 (this issue).
12. Gehrels, N., Ramirez-Ruiz, E. & Fox, D. B. Gamma-ray bursts in the Swift Era. *Annu. Rev. Astron. Astrophys.* **47**, 567–617 (2009).
13. Campana, S. et al. The association of GRB 060218 with a supernova and the evolution of the shock wave. *Nature* **442**, 1008–1010 (2006).
14. Chornock, R., Marion, G. H., Narayan, G., Berger, E. & Soderberg, A. M. GRB101225A: MMT spectroscopy. *GCN Circ.* **11507** (2010).

15. Tanvir, N. R. *et al.* GRB101225A: HST observations - no host detected. *GCN Circ.* **11564** (2011).
16. Gänsicke, B. T. *et al.* A gaseous metal disk around a white dwarf. *Science* **314**, 1908–1910 (2006).
17. Jura, M. A tidally disrupted asteroid around the white dwarf G29-38. *Astrophys. J.* **584**, L91–L94 (2003).
18. Ulmer, A. Flares from the tidal disruption of stars by massive black holes. *Astrophys. J.* **514**, 180–187 (1999).
19. Cannizzo, J. K., Lee, H. M. & Goodman, J. The disk accretion of a tidally disrupted star onto a massive black hole. *Astrophys. J.* **351**, 38–46 (1990).
20. Burrows, D. N. *et al.* Onset of a relativistic jet from the tidal disruption of a star by a massive black hole. *Nature* **476**, 421–424 (2011).
21. Lodato, G. & Rossi, E. M. Multiband light curves of tidal disruption events. *Mon. Not. R. Astron. Soc.* **410**, 359–367 (2011).
22. Guillochon, J., Ramirez-Ruiz, E. & Lin, D. Consequences of the ejection and disruption of giant planets. *Astrophys. J.* **732**, 74 (2011).
23. Lasota, J.-P. The disc instability model of dwarf novae and low-mass X-ray binary transients. *N. Astron. Rev.* **45**, 449–508 (2001).
24. Lasota, J.-P., Dubus, G. & Kruk, K. Stability of helium accretion discs in ultracompact binaries. *Astron. Astrophys.* **486**, 523–528 (2008).
25. Shull, J. M. & Stern, S. A. Gamma-ray burst constraints on the galactic frequency of extrasolar Oort Clouds. *Astron. J.* **109**, 690–697 (1995).
26. Jura, M. An upper bound to the space density of interstellar comets. *Astron. J.* **141**, 155 (2011).
27. Wolszczan, A. & Frail, D. A. A planetary system around the millisecond pulsar PSR1257+12. *Nature* **355**, 145–147 (1992).
28. Sigurdsson, S., Richer, H. B., Hansen, B. M., Stairs, I. H. & Thorsett, S. E. A young white dwarf companion to pulsar B1620-26: evidence for early planet formation. *Science* **301**, 193–196 (2003).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements The authors acknowledge support from ASI and INAF. S. Campana wants to thank A. Possenti and N. Tanvir for conversations, and acknowledges N. Scharrel for granting a DDT XMM-Newton observation.

Author Contributions S. Campana led the research and wrote the majority of manuscript, invented the tidal disruption model, and did the X-ray data analysis. G.L. and E.M.R. worked on the tidal disruption model, fitted the light curves providing the minor-body parameters, and estimated the rate of events. P.D.'A. analysed optical data. N.P., M.D.V., G.T. and S. Covino contributed to the exclusion of Galactic models. G. Ghisellini, G. Ghirlanda, N.P., E. Pian and M.D.V. contributed to the exclusion of extragalactic models. R.S. contributed to exclude the possibility of tidal disruption onto an intermediate-mass black hole. L.A.A., A.M., G.C., V.D.'E., D.F., E. Palazzi, B.S. and S.D.V. contributed in obtaining optical data and provided an unbiased reading of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to S. Campana (sergio.campana@brera.inaf.it).

The unusual γ -ray burst GRB 101225A from a helium star/neutron star merger at redshift 0.33

C. C. Thöne^{1,2}, A. de Ugarte Postigo³, C. L. Fryer⁴, K. L. Page⁵, J. Gorosabel¹, M. A. Aloy⁶, D. A. Perley⁷, C. Kouveliotou⁸, H. T. Janka⁹, P. Mimica⁶, J. L. Racusin¹⁰, H. Krimm^{10,11,12}, J. Cummings¹⁰, S. R. Oates¹³, S. T. Holland^{10,11,12}, M. H. Siegel¹⁴, M. De Pasquale¹³, E. Sonbas^{10,11,15}, M. Im¹⁶, W.-K. Park¹⁶, D. A. Kann¹⁷, S. Guziy^{1,18}, L. Hernández García¹, A. Llorente¹⁹, K. Bundy⁷, C. Choi¹⁶, H. Jeong²⁰, H. Korhonen^{21,22}, P. Kubànek^{1,23}, J. Lim²⁴, A. Moskvitin²⁵, T. Muñoz-Darias²⁶, S. Pak²⁰ & I. Parrish⁷

Long γ -ray bursts (GRBs) are the most dramatic examples of massive stellar deaths, often associated with supernovae¹. They release ultra-relativistic jets, which produce non-thermal emission through synchrotron radiation as they interact with the surrounding medium². Here we report observations of the unusual GRB 101225A. Its γ -ray emission was exceptionally long-lived and was followed by a bright X-ray transient with a hot thermal component and an unusual optical counterpart. During the first 10 days, the optical emission evolved as an expanding, cooling black body, after which an additional component, consistent with a faint supernova, emerged. We estimate its redshift to be $z = 0.33$ by fitting the spectral-energy distribution and light curve of the optical emission with a GRB-supernova template. Deep optical observations may have revealed a faint, unresolved host galaxy. Our proposed progenitor is a merger of a helium star with a neutron star that underwent a common envelope phase, expelling its hydrogen envelope. The resulting explosion created a GRB-like jet which became thermalized by interacting with the dense, previously ejected material, thus creating the observed black body, until finally the emission from the supernova dominated. An alternative explanation is a minor body falling onto a neutron star in the Galaxy³.

On 25 December 2010, at 18:37:45 UT, the Burst Alert Telescope (BAT, 15–350 keV) on board the Swift satellite detected GRB 101225A, one of the longest GRBs ever observed by Swift⁴ (see Supplementary Information); this GRB had $T_{90} > 2,000$ s (T_{90} is the time in which 90% of the γ -ray energy is released⁵). A bright X-ray afterglow was detected for two days, and a counterpart in the ultraviolet, optical and infrared bands could be observed from 0.38 hours to two months after the event (see Supplementary Information). No counterpart was detected at radio frequencies^{6,7}.

The most surprising feature of GRB 101225A is the spectral energy distribution (SED) of its afterglow. The X-ray SED is best modelled with a combination of an absorbed power-law and a black body. The ultraviolet/optical/near-infrared (UVOIR) SED (see Fig. 1) can be fitted with a cooling and expanding black-body model until 10 days after the burst (see Supplementary Information), after which we observe an additional spectral component accompanied by a flattening of the light curve (Fig. 2). This behaviour differs from a normal GRB where

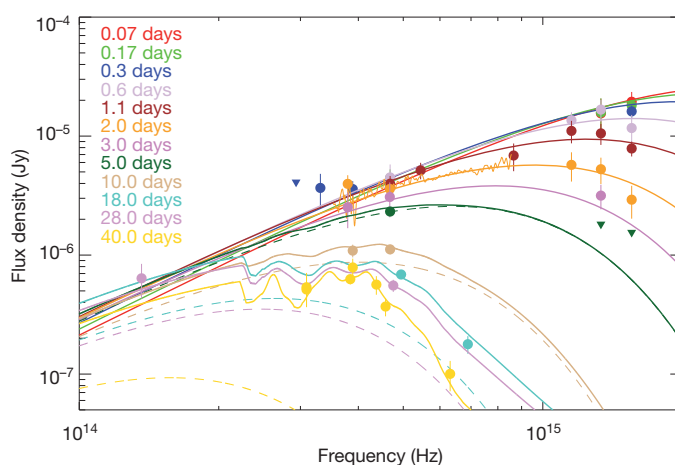


Figure 1 | Temporal evolution of the ultraviolet, optical and infrared (UVOIR) spectral energy distribution. The ultraviolet, optical and infrared counterparts were detected by UVOT (the ultraviolet telescope on board Swift) and several ground-based facilities, from 0.38 h to nearly 2 months after the GRB. This plot shows the evolution of the SED from the onset of the optical observations at 0.07 days to 40 days for all epochs with sufficient data to model the SED shape. Filled circles, detections; triangles, upper limits; error bars, 1σ . The additional orange line on top of the smooth model at 2.0 days shows our flux-calibrated spectrum taken with the OSIRIS/GTC. The SED evolution requires two different components, a simple expanding and cooling black body up to ~ 10 days and an additional supernova component for the last four epochs. The solid lines show the combined evolution of the black body and supernova contributions, the dashed lines from day 5 on show the evolution of the black body component alone. The UVOIR black body evolves from an initial temperature of 43,000 K (0.07 d) to 5,000 K (18 d) and increases in radius from 2×10^{14} cm to 7×10^{14} cm at the same timescale. We used the SED at 40 days to fit the supernova component with a template of the broad-line type Ic supernova 1998bw which was associated with GRB 980425. Reanalysing UVOIR data of XRF 060218¹³ and SN 2008D²², we find a similar thermal component over the first 3–4 days, but with an earlier onset of the supernova component (see Supplementary Information).

¹IAA – CSIC, Glorieta de la Astronomía s/n, 18008 Granada, Spain. ²Niels Bohr International Academy, Niels Bohr Institute, Blegdamsvej 17, 2100 Copenhagen, Denmark. ³Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen, Denmark. ⁴Los Alamos National Laboratory, MS D409, CCS-2, Los Alamos, New Mexico 87545, USA. ⁵Department of Physics and Astronomy, University of Leicester, University Road, Leicester LE1 7RH, UK. ⁶Departamento de Astronomía y Astrofísica, Universidad de Valencia, 46100 Burjassot, Spain. ⁷Astronomy Department, UC Berkeley, 601 Campbell Hall, Berkeley, California 94720, USA. ⁸Science and Technology Office, ZP12, NASA/Marshall Space Flight Center, Huntsville, Alabama 35812, USA. ⁹Max-Planck-Institut für Astrophysik, Karl-Schwarzschild-Strasse 1, 85748 Garching, Germany. ¹⁰NASA, Goddard Space Flight Center, Greenbelt, Maryland 20771, USA. ¹¹Universities Space Research Association, 10211 Wincopin Circle, Suite 500, Columbia, Maryland 21044-3432, USA. ¹²Center for Research and Exploration in Space Science and Technology (CREST), 10211 Wincopin Circle, Suite 500, Columbia, Maryland 21044-3432, USA. ¹³Mullard Space Science Laboratory, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK. ¹⁴Department of Astronomy and Astrophysics, Pennsylvania State University, 104 Davey Laboratory, University Park, Pennsylvania 16802, USA. ¹⁵University of Adiyaman, Department of Physics, 02040 Adiyaman, Turkey. ¹⁶Center for the Exploration of the Origin of the Universe, Department of Physics and Astronomy, Seoul National University, 56-1 San, Shillim-dong, Kwanak-gu, Seoul, South Korea. ¹⁷Thüringer Landessternwarte Tautenburg, Sternwarte 5, 07778 Tautenburg, Germany. ¹⁸Nikolaev National University, Nikolska 24, Nikolaev, 54030, Ukraine. ¹⁹Herschel Science Operations Centre, INSA, ESAC, Villafranca del Castillo, PO Box 50727, I-28080 Madrid, Spain. ²⁰School of Space Research, Kyung Hee University, 1 Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do 446-701, South Korea. ²¹Finnish Centre for Astronomy with ESO (FINCA), University of Turku, Väisälantie 20, 21500 Piikkiö, Finland. ²²Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, 2100 Copenhagen, Denmark. ²³Institute of Physics, Na Slovance 2, 180 00, Prague 8, Czech Republic. ²⁴Department of Astronomy and Space Science, Kyung Hee University, 1 Seocheon-dong, Giheung-gu, Yongin-si, Gyeonggi-do 446-701, Korea. ²⁵Special Astrophysical Observatory of the Russian Academy of Sciences, Nizhny Arkhyz 369167, Russia. ²⁶INAF – Osservatorio Astronomico di Brera, Via E. Bianchi 46, 23807 Merate, Italy.

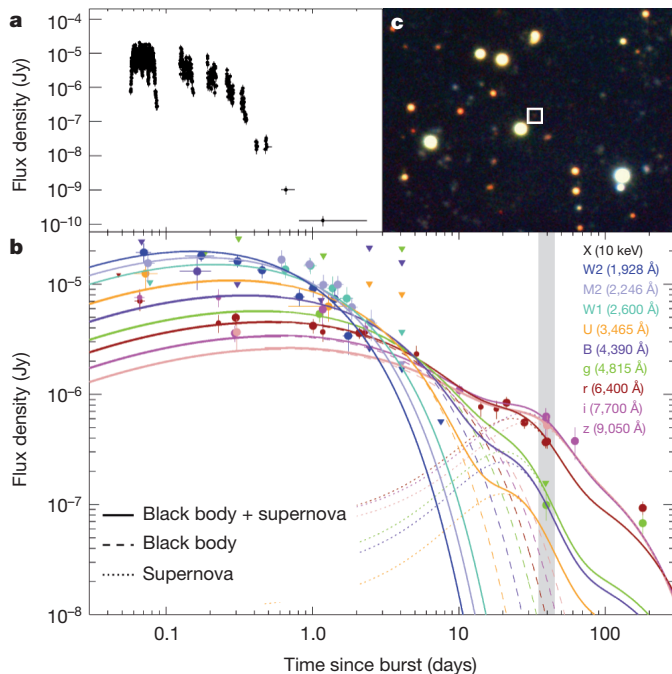


Figure 2 | Light curves of GRB 101225A in X-ray and ultraviolet/optical/infrared bands. **a**, X-ray light curve; **b**, light curves in ultraviolet to infrared bands (colour coded by wavelength; see key). Filled circles, detections; large symbols, our observations; small symbols, values from the literature; triangles, upper limits; error bars, 1σ . Solid lines are the combined light curve from the black-body and the supernova component, excluding the contribution from the host galaxy. Dashed lines and dotted lines show the evolution of the black-body component and of the supernova, respectively. Observations started almost simultaneously in X-rays and optical/ultraviolet wavelengths. The X-rays reached a peak flux of $4.34 \times 10^{-9} \text{ erg cm}^{-2} \text{ s}^{-1}$. After an initial shallow decay of slope $t^{-1.108 \pm 0.011}$ up to 21 ks, the X-rays show a strong decay with a slope of $t^{-5.95 \pm 0.20}$, inconsistent with synchrotron emission. The UVOIR light curves have a shallow maximum at the beginning, with different peak times for the individual bands due to the maximum of the black-body emission passing through the spectrum. The second component emerging at around 10 days post-burst is the contribution of an underlying supernova, modelled with the GRB-SN 1998bw as a template, stretched in time by a factor of 1.25 and decreased in luminosity by a factor of 12 (in restframe). The absolute luminosity of the supernova is $M_{V,\text{abs}} = -16.7$ mag, the faintest supernova associated with a GRB. At ~ 180 days we detect the very faint host at magnitudes of $g' = 27.36 \pm 0.27$ and $r' = 26.90 \pm 0.14$ or $M_{g,\text{abs}} = -13.7$ ($0.001 L^*$, where L^* is the characteristic luminosity). **c**, Colour image of the field of GRB 101225A observed at 40 days (indicated by a grey bar in **b**) with the afterglow location marked by a box.

the SED follows a power law owing to synchrotron emission created in shocks when the jet hits the interstellar medium (see, for example, ref. 2).

An optical spectrum taken two nights after the burst does not show any spectral lines (see Supplementary Information). We fitted the SED and light curve with the template of SN 1998bw, a type Ic supernova associated with GRB 980425⁸, and obtained a redshift of $z = 0.33$ (see Supplementary Information). At this distance, the supernova has an absolute peak magnitude of only $M_{V,\text{abs}} = -16.7$ mag, which makes it the faintest supernova associated with a long GRB^{9,10}. In contrast, the γ -ray isotropic-equivalent energy release at $z = 0.33$ is $> 1.4 \times 10^{51}$ erg, typical of other long GRBs but more luminous than most other low-redshift GRBs associated with supernovae¹¹. We detect a possible host galaxy in g' and r' bands with the OSIRIS instrument on GTC (Gran Telescopio Canarias) at 6 months after the burst with an absolute magnitude of only $M_{g,\text{abs}} = -13.7$ mag, ~ 2 mag fainter than any other GRB host¹². Although its blue colour matches that of a star-forming galaxy, our observations do not allow us to resolve it as an extended source.

At $z = 0.33$, the X-ray black body has a radius of $\sim 2 \times 10^{11}$ cm (~ 3 solar radii) and a temperature of ~ 1 keV (10^7 K) at 0.07 d with little temporal evolution. Such a thermal component, attributed to the shock breakout from the star, has also been observed for XRF 060218¹³, XRF 100316D/SN 2010dh¹⁴ and GRB 090618¹⁵, all nearby GRBs associated with type Ic supernovae^{14,16,17,18}, with similar temperatures but larger radii. The UVOIR black body starts with a radius of 2×10^{14} cm (~ 13 AU) and a temperature of 8.5×10^4 K at similar times and evolves considerably over the next 10 days, reaching a radius of 7×10^{14} cm and temperature of 5,000 K. The evolution of the two black-body components suggests that they must stem from different processes and regions (see Supplementary Information).

An appealing model is a helium star–neutron star merger with a common envelope phase, a model that has been proposed earlier as a possible progenitor for GRBs^{19–21}. In this scenario, a binary system consisting of two massive stars survives the collapse of the more massive component to a neutron star. When the second star leaves the main sequence and expands, it engulfs the neutron star, leading to a common-envelope phase and the ejection of the hydrogen envelope and part of the helium core as the remnant spirals into the centre of the second star. When the neutron star reaches the centre, angular momentum forms a disk around the remnant of the merger, allowing for the formation of a GRB-like jet. This remnant might be a magnetar whose prolonged activity can explain the very long duration of the GRB.

The interaction of this ultra-relativistic, well-collimated jet with the previously ejected common-envelope material can explain both the X-ray and UVOIR emission components. Estimating that the in-spiral takes 5 orbits or 1.5 yr and material is ejected at escape velocity, the outer ejecta are at a distance of a few times 10^{14} cm at the time of the merger, consistent with the radius of the UVOIR black body. We assume that the ejecta form a broad torus with a narrow, low-density funnel along the rotation axis of the system that permits the passage of the γ -radiation generated in the jet. Most of the jet hits the inner boundary of the common-envelope ejecta and only a small fraction of it propagates through the funnel. The X-ray emission is produced by shocks created by the interaction of the jet with the inner boundary of the common-envelope shell. As the jet passes through the funnel, it decelerates owing to the increased baryon load and shear with the funnel walls so that a regular afterglow signature is suppressed. When the now mildly relativistic, mass-loaded jet breaks out of the common-envelope ejecta, it produces the UVOIR emission in the first 10 d. As the supernova shock expands beyond the common-envelope shell, we observe a small bump in the light curve at ~ 30 d. This helium star/neutron star merger scenario naturally assumes the production of a relatively small amount of radioactive nickel, leading to a weak supernova (for a detailed description of the different processes, see Supplementary Information).

A similar scenario might explain another, previously detected, event, XRF 060218¹³, which showed a thermal component both in X-rays and at optical wavelengths (see Supplementary Information), albeit with a different progenitor system producing a brighter supernova and a fainter GRB. On the other hand, a class of GRBs exist that show a thermal component in X-rays, but have a classical afterglow with a power-law SED, such as GRB 090618¹⁵. Finally, SN 2008D, a type Ib supernova in NGC 2770²² showing X-ray emission, had an early thermal component in the optical emission (see Supplementary Information), attributed to the shock breakout and independent of the supernova emission itself. GRB 101225A might hence be, together with XRF 060218, a member of a newly defined class of ‘black-body-dominated’, supernova-associated, long-duration GRBs, which arise in very dense environments that are created by the progenitor systems themselves; this dense environment thermalizes the high-energy output from the collapsing star. The non-relativistic, uncollimated emission in this scenario makes it difficult to detect such events at higher redshifts. This makes GRB 101225A a fortunate case that allows us to derive

conclusions about the progenitor system and its environment from a new variety of massive stellar death, which had so far been only proposed to exist theoretically.

Received 10 April; accepted 3 October 2011.

1. Woosley, S. E. & Bloom, J. S. The supernova gamma-ray burst connection. *Annu. Rev. Astron. Astrophys.* **44**, 507–556 (2006).
2. Zhang, B. & Mészáros, P. Gamma-ray bursts: progress, problems and prospects. *Int. J. Mod. Phys. A* **19**, 2385–2472 (2004).
3. Campana, S. *et al.* The unusual gamma-ray burst GRB 101225A explained as a minor body falling onto a neutron star. *Nature* doi:10.1038/nature10592 (this issue).
4. Sakamoto, T. *et al.* The second Swift Burst Alert Telescope gamma-ray burst catalog. *Astrophys. J. Suppl. Ser.* **195**, 2 (2011).
5. Kouveliotou, C. *et al.* Identification of two classes of gamma-ray bursts. *Astrophys. J.* **413**, L101–L104 (1993).
6. Frail, D. A. GRB 101225A. *GCN Circ.* **11550** (2011).
7. Zauderer, A., Berger, E. & Fong, W. Early EVLA observations of Christmas burst. *GCN Circ.* **11770** (2011).
8. Galama, T. J. *et al.* An unusual supernova in the error box of the γ -ray burst of 25 April 1998. *Nature* **395**, 670–672 (1998).
9. Soderberg, A. *et al.* An HST study of the supernovae accompanying GRB 040924 and GRB 041006. *Astrophys. J.* **636**, 391–399 (2006).
10. Wiersema, K. *et al.* Spectroscopy and multiband photometry of the afterglow of intermediate duration γ -ray burst GRB 040924 and its host galaxy. *Astron. Astrophys.* **481**, 319–326 (2008).
11. Kann, D. A. *et al.* The afterglows of Swift-era gamma-ray bursts. I. Comparing pre-Swift and Swift-era long/soft (type II) GRB optical afterglows. *Astrophys. J.* **720**, 1513–1558 (2010).
12. Wiersema, K. *et al.* The nature of the dwarf star-forming galaxy associated with GRB 060218/SN 2006aj. *Astron. Astrophys.* **464**, 529–539 (2007).
13. Campana, S. *et al.* The association of GRB 060218 with a supernova and the evolution of the shock wave. *Nature* **442**, 1008–1010 (2006).
14. Starling, R. L. C. *et al.* Discovery of the nearby long, soft GRB 100316D with an associated supernova. *Mon. Not. R. Astron. Soc.* **411**, 2792–2803 (2011).
15. Page, K. L. *et al.* GRB 090618: detection of thermal X-ray emission from a bright gamma-ray burst. *Mon. Not. R. Astron. Soc.* **416**, 2078–2089 (2011).
16. Pian, E. *et al.* An optical supernova associated with the X-ray flash XRF 060218. *Nature* **442**, 1011–1013 (2006).
17. Mazzali, P. *et al.* A neutron-star-driven X-ray flash associated with supernova SN 2006aj. *Nature* **442**, 1018–1020 (2006).
18. Cano, Z. *et al.* A tale of two GRB-SNe at a common redshift of $z=0.54$. *Mon. Not. R. Astron. Soc.* **413**, 669–685 (2011).
19. Fryer, C. L. & Woosley, S. E. Helium star/black hole mergers: a new gamma-ray burst model. *Astrophys. J.* **502**, L9–L12 (1998).
20. Zhang, W. & Fryer, C. L. The merger of a helium star and a black hole: gamma-ray bursts. *Astrophys. J.* **550**, 357–367 (2001).
21. Barkov, M. V. & Komissarov, S. S. Recycling of neutron stars in common envelopes and hypernova explosions. *Mon. Not. R. Astron. Soc.* **415**, 944–958 (2011).
22. Soderberg, A. *et al.* An extremely luminous X-ray outburst at the birth of a supernova. *Nature* **453**, 469–474 (2008).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This Letter is based on observations collected at CAHA/Calar Alto, GTC/La Palma, the Liverpool Telescope at ORM/La Palma, the McDonald Observatory at the University of Texas at Austin, and Gemini-North and Keck on Hawaii. We thank J. S. Bloom for helping with the Keck observations. The Dark Cosmology Centre is funded by the DNRf. K.L.P., S.R.O. and M.D.P. acknowledge the support of the UK Space Agency. J.G., S.G. and P.K. are partially supported by MICINN. M.A.A. and P.M. are supported by an ERC starting grant. H.T.J. acknowledges support by a DFG grant. M.I., W.-K.P., C.C., J.L. and S.P. acknowledge support from CRI/NRF/MEST of Korea. A.M. acknowledges support from the Russian government.

Author Contributions C.C.T. did the overall management of the observations and modelling, the analysis of the spectra and wrote most of the manuscript. A.d.U.P. did the UVOIR black-body modelling, supernova template fitting, most of the optical/infrared photometry and lead the GTC observations. C.L.F. suggested and investigated the progenitor system. K.L.P. did the X-ray analysis, J.G. worked on the supernova templates and the photometric calibrations for the optical/infrared data. M.A.A. did the modelling of the UVOIR black body and X-ray emission from numerical simulations. D.A.P. contributed to the observation and analysis of the late Gemini and Keck data. C.K. investigated possible progenitor models. H.T.J., P.M. and A.L. contributed to the theoretical modelling. J.L.R., H.K., J.C., S.R.O., S.T.H., M.H.S., M.D.P. and E.S. did the analysis of the Swift data. M.I., W.-K.P., C.C., H.J., J.L. and S.P. contributed the McDonald 2.1-m data, A.M. the late BTA 6-m data, K.B. and I.P. the late Keck spectrum. D.A.K. did the comparison of supernova stretching factors and luminosities. S.G. and L.H.G. helped with the optical photometry. H.K. and T.M.-D. investigated alternative interpretations of the event, and P.K. assisted with the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to C.C.T. (cchoene@iaa.es).

Observation of a pairing pseudogap in a two-dimensional Fermi gas

Michael Feld^{1*}, Bernd Fröhlich^{1*}, Enrico Vogt¹, Marco Koschorreck¹ & Michael Köhl¹

Pairing of fermions is ubiquitous in nature, underlying many phenomena. Examples include superconductivity, superfluidity of ^3He , the anomalous rotation of neutron stars, and the crossover between Bose–Einstein condensation of dimers and the BCS (Bardeen, Cooper and Schrieffer) regime in strongly interacting Fermi gases. When confined to two dimensions, interacting many-body systems show even more subtle effects¹, many of which are not understood at a fundamental level. Most striking is the (as yet unexplained) phenomenon of high-temperature superconductivity in copper oxides, which is intimately related to the two-dimensional geometry of the crystal structure. In particular, it is not understood how the many-body pairing is established at high temperature, and whether it precedes superconductivity. Here we report the observation of a many-body pairing gap above the superfluid transition temperature in a harmonically trapped, two-dimensional atomic Fermi gas in the regime of strong coupling. Our measurements of the spectral function of the gas are performed using momentum-resolved photoemission spectroscopy^{2,3}, analogous to angle-resolved photoemission spectroscopy in the solid state⁴. Our observations mark a significant step in the emulation of layered two-dimensional strongly correlated superconductors using ultracold atomic gases.

One essential mechanism to establish collective quantum phases in Fermi systems, such as superconductivity or superfluidity, is fermionic pairing. In the case of weak attractive interactions between spin-up and spin-down fermions in three dimensions, Cooper pairs form on the surface of the Fermi sea, even though in vacuum no two-particle bound state exists. Pairing and the condensation of pairs go hand in hand. However, in this weak coupling limit, the gap and hence the superconducting transition temperature T_c are exponentially small. The fundamental and technological quest for increasing the superconducting transition temperature has led to the discovery of correlated materials that depart from weak coupling theory. These are complex compounds that combine strong interactions and large quantum fluctuations (owing to a two-dimensional geometry), both of which hinder a precise understanding and thorough theoretical modelling, even at zero temperature. Strongly correlated, two-dimensional, layered superconductors have displayed a wealth of new and unexpected phenomena. Among these is the peculiar observation of a suppression of low-energy weight in the single-particle spectral function above T_c and hence without direct link to superconducting behaviour^{5–7}. The origin of this pseudogap phenomenon so far lacks full understanding, and the implications of the pseudogap for the appearance of superconductivity have remained an open question.

Ultracold atomic gases provide an ideal testing ground to uncover the physics of strongly correlated Fermi systems because of the exceptional capability to tune the system parameters. Experiments in three-dimensional Fermi gases have observed fermionic many-body pairing in the crossover regime from BCS-like pairing to Bose–Einstein condensation of dimers (for a review, see ref. 8). Neither in the BCS-limit nor in the Bose–Einstein condensate (BEC)-limit is a pseudogap

expected in three dimensions. In between, that is, for $k_F|a| \gg 1$, the gas is in the unitary regime of strongest interactions. Here k_F is the Fermi wave vector, and a is the s -wave scattering length between spin-up and spin-down particles. Whether a pairing pseudogap phase exists here or whether a Fermi-liquid state dominates is still debated^{9–17}.

When the motion of particles is constrained to a two-dimensional plane, a bound state with binding energy E_B exists for any value of the three-dimensional scattering length^{8,18,19}. This is in contrast to the situation in three dimensions, where weakly bound molecules start to appear in the unitary regime of strongest interaction. The presence of the two-body bound state in vacuum is a necessary and sufficient condition for many-body pairing in a medium in two dimensions²⁰, and a small binding energy $E_B \ll E_F$ corresponds to the attractive BCS-like regime with interaction parameter $\ln(k_F a_{2D}) > 1$. Here E_F denotes the Fermi energy, $a_{2D} = \hbar/\sqrt{mE_B} > 0$ is the two-dimensional scattering length, m is the mass of the atoms and \hbar is Planck's constant divided by 2π . At zero temperature, BCS-type superfluidity has been predicted in the attractive two-dimensional mean-field regime^{20–23}.

Qualitatively, we understand the behaviour at strong coupling and finite temperature by considering a complex order parameter¹, $\Delta(\mathbf{x}) = |\Delta(\mathbf{x})|e^{i\theta(\mathbf{x})}$. The phase $\theta(\mathbf{x})$ may fluctuate as a function of the spatial coordinate \mathbf{x} , for example, driven by thermal fluctuations and quantum fluctuations enhanced in low-dimensional systems. For temperatures below the Berezinskii–Kosterlitz–Thouless transition temperature T_{BKT} , the superfluid gap $\Delta_{\text{sc}} = \langle \Delta \rangle$ is larger than zero. For higher temperatures, $T > T_{\text{BKT}}$, phase fluctuations destroy long range order (that is, $\langle e^{i\theta(\mathbf{x})} \rangle = 0$), however, the modulus $\langle |\Delta| \rangle$ remains finite up to the pairing crossover temperature T^* , which is of the order of E_B . The regime between T_{BKT} and T^* is referred to as the pseudogap regime, and is still little understood. The transition between the normal phase and the pseudogap phase is not associated with breaking a continuous symmetry but is a crossover phenomenon. This scenario is different from standard weak-coupling mean-field BCS theory in which the superfluid gap is destroyed by pair breaking and T_c and T^* are approximately equal. The temperature range for accessing the pseudogap phase, that is, the difference between T^* and T_{BKT} , is particularly large in two dimensions, because enhanced quantum fluctuations and second-order interaction effects²⁴ suppress T_{BKT} with respect to the mean-field result. The pseudogap regime is connected by a crossover to a Bose-liquid regime of local pairs when $\ln(k_F a_{2D})$ crosses zero or, equivalently, $E_B = 2E_F$.

Here, we study two-dimensional Fermi gases for $T > T_{\text{BKT}}$ in the strongly interacting regime $|\ln(k_F a_{2D})| \leq 1$, in which the size of the pairs is comparable to the mean interparticle spacing. Hence, Pauli blocking and other many-body effects act on the pairing, and a key question is whether a pseudogap phase exists above the superfluid transition temperature^{25,26}, $T_{\text{BKT}} \approx 0.1T_F$. We investigate many-body pairing using momentum-resolved radio-frequency (r.f.) spectroscopy (see Methods). To this end, a long-wavelength photon of energy $\hbar\Omega$ creates a single-particle excitation in the two-dimensional gas, which we detect at momentum k . This measures the spectral function $A(k, \Omega)$

¹Cavendish Laboratory, University of Cambridge, JJ Thomson Avenue, Cambridge CB3 0HE, UK.

*These authors contributed equally to this work.

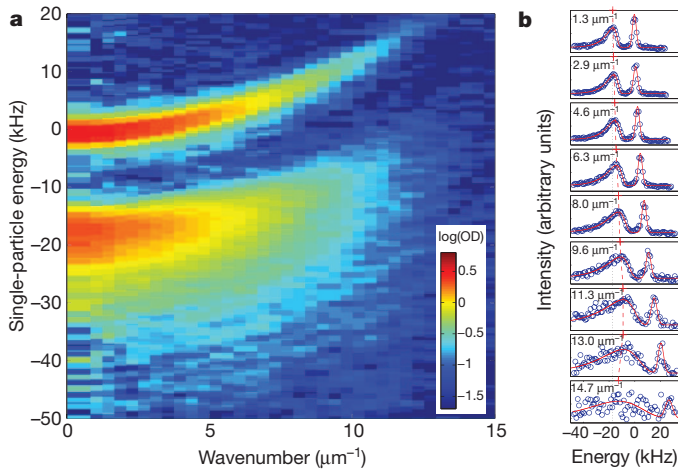


Figure 1 | Measuring the spectral function. **a**, Measured photoemission signal at $\ln(k_F a_{2D}) = 0$. The image is averaged over six repetitions, and we plot \log_{10} of the optical density (OD). The signal is proportional to the single-particle spectral function $A(k, \Omega)$. **b**, Energy distribution curves for various values of the momentum, $\hbar k$ (shown top left in each subpanel). The data of the pairing peak (lower feature in **a**) are fitted with a modified Gumbel distribution (solid red curve) and the peak positions (red crosses, connected by red dashed line) indicate the back-bending of the dispersion curve as compared to the $k = 0$ peak position (dotted line). The data of the atomic peak are fitted with a Gaussian.

multiplied by the fermionic occupation function. This technique was originally developed for the study of solid state materials, such as copper oxide superconductors⁴, but recently has been successfully adapted to cold atomic systems^{2,3}.

Figure 1a shows a measured spectral function, typical of the data sets we use in this Letter. The spectrum exhibits two features: the upper branch near zero energy corresponds to unpaired atoms, most probably from low-density regions of the trap, with vanishingly small interaction between the free atoms²⁷. The lower branch corresponds to the pairing signal with a non-trivial dispersion in the spectral function. In Fig. 1b we plot the spectra $A(\Omega)$ for constant values of k , which show clearly asymmetric line shapes as expected for pair breaking into a continuum, together with our fits (see Methods).

We analyse the spectral function for various interaction strengths and temperatures. In the attractively interacting regime $\ln(k_F a_{2D}) > 0$ the size a_{2D} of the pairs is larger than the mean interparticle spacing k_F^{-1} . Therefore, at low temperature no isolated dimers exist in the

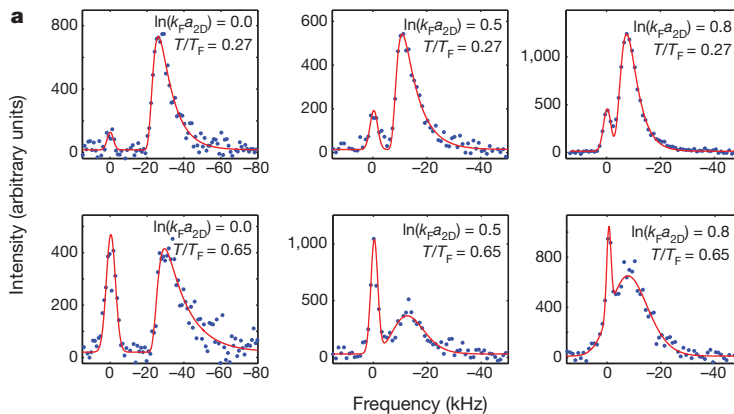
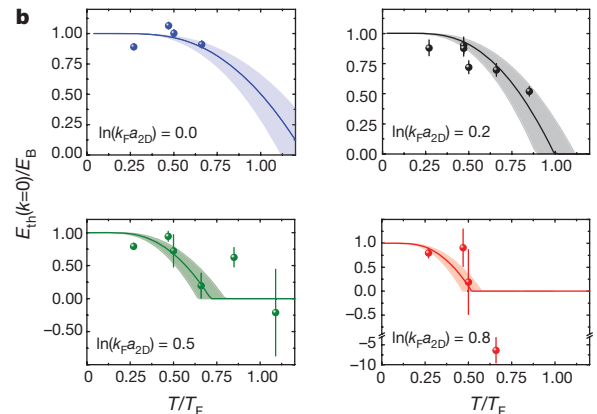


Figure 2 | Pairing in the strongly interacting two-dimensional Fermi gas. **a**, Energy distribution curves extracted from the photoemission signal at $k = 0$ (proportional to $A(k = 0, \Omega)$) for different values of $\ln(k_F a_{2D})$ and temperatures T/T_F (shown top right in each subpanel). For low temperatures, hardly any unpaired atoms exist. The solid line is the fit described in the text. **b**, Energy threshold of the pairing peak as a function of temperature for different

sample but a many-body pairing gap opens in the spectrum, conceptually similar to a BCS-paired state in three dimensions. The zero-temperature limit of the two-dimensional BEC-BCS crossover has been investigated in the framework of mean-field theory^{1,20}. This has led to the remarkable prediction that at zero temperature the condensation energy per particle $E_0 = \Delta_{sc}^2/2E_F$ at $k = 0$ should be equal to the two-body binding energy in vacuum E_B for all interaction strengths.

The expected r.f. spectrum can be qualitatively estimated on the basis of mean-field theory of a BCS state at zero temperature. Many-body pairing causes an energy threshold E_{th} , and this energy has to be supplied by the r.f. field to create a single-particle excitation. Ignoring final state interactions and assuming two-dimensional geometry with a constant density of states, the momentum integrated lineshape has the form $\Gamma(\Omega) \propto \theta(\hbar\Omega - E_{th})/\Omega^2$, where $\theta(x)$ is the Heaviside function. We investigate many-body pairing at low temperatures by studying the energy threshold E_{th} of the pairing peak in the photoemission spectrum at momentum $k = 0$. This quantity is closely linked to E_0 , and its disappearance, for example as temperature is increased, indicates the transition into a normal phase. Investigating E_{th} at $k = 0$, rather than for example at k_F , has the advantage that averaging over the inhomogeneous spread of k_F and the chemical potential μ of the approximately 30 two-dimensional gases (see Methods) has a small effect because the mean-field prediction at zero temperature is not explicitly dependent on μ and k_F . At the temperature $T/T_F = 0.27$ (see Fig. 2a, top row), we observe a negligible contribution of free atoms and a sharp onset of the pairing peak. As we increase the temperature of our sample (see Fig. 2a, bottom row), the spectral weight at low energies increases because the pairing peak broadens and the threshold shifts. For $\ln(k_F a_{2D}) = 0.8$ at $T/T_F = 0.65$, for example, the broadening creates a significant amount of low-energy spectral weight and the threshold of the spectrum even moves to positive energies. The latter result shows that the energy gap has completely disappeared.

In Fig. 2b we show the measured temperature dependence of E_{th} for various interaction strengths on the BCS side of the resonance $\ln(k_F a_{2D}) > 0$, where no local dimers exist in the sample. We observe that E_{th} moves towards zero with increasing temperature, which indicates that the many-body pairing gap vanishes. We compare our data to finite temperature mean-field theory in two dimensions¹. Overall, the numerically calculated energy $E_0 = |\Delta(T)|^2/2E_F$ reproduces our experimental data for E_{th} without free parameters (solid lines in Fig. 2b). However, some discrepancies remain. For example, for the coldest temperatures the average of E_{th}/E_B over all interaction strengths is 0.84 ± 0.05 , approximately 15% smaller than predicted



interaction strengths (shown bottom left in each subpanel). The solid lines are predictions of mean-field theory¹ for the quantity $|\Delta(T)|^2/2E_F$ in two dimensions without free parameters. The shaded area marks the uncertainty of our determination of $\ln(k_F a_{2D})$, which is ± 0.1 . Error bars, 1σ fit errors. They become larger towards larger values of $\ln(k_F a_{2D})$ because of the division by the exponentially smaller binding energy E_B .

by mean-field theory. The deviation could stem from beyond mean-field effects provoked by our two-dimensional geometry and interaction energy shifts. It could be more precisely computed using quantum Monte Carlo calculations, which, however, up to now exist only for zero temperature²³. From our data we conclude that we have realized a many-body pairing phase in two dimensions and observed how it vanishes as the temperature is increased to approach the pairing crossover temperature, T^* . Our data are taken above the critical temperature for superfluidity T_{BKT} in two dimensions, which is predicted to be at $T/T_F \approx 0.1$ in the strong coupling regime. We have verified this experimentally by using a rapid magnetic field sweep across the Feshbach resonance to project the possibly condensed pairs onto deeply bound molecules and then image the molecular condensate after time-of-flight²⁸. Whereas this procedure allows us to observe three-dimensional Fermi condensates, in our two-dimensional samples no condensation was detected. We therefore identify our phase as a pairing pseudogap phase.

In addition to the energy threshold at zero momentum, we also investigate how the quasiparticle dispersion is affected when changing temperature and interaction strength. We extract the dispersion data from the peak of the spectral function $E_{\text{peak}}(k) = \max_{\Omega} [A(k, \Omega)]$ (see Fig. 1b). In the crossover region and on the BEC side of the resonance, the dispersion shows a pronounced concave shape at large momenta ('back-bending'), which we display in Fig. 3a for various interaction strengths. This shape is reminiscent of the hole dispersion of a BCS state and qualitatively similar to that of a three-dimensional Fermi gas near unitarity³. We fit the dispersion relation $E(k) = \mu - \sqrt{[\hbar^2/(2m^*)(k^2 - k^{*2})]^2 + \Delta^2}$ for $k \leq 1.5k_F$ using the zero temperature mean-field prediction for the chemical potential²⁰ to extract the parameters effective mass m^* and the wave vector k^* , at which the back-bending occurs. In Fig. 3b and c we display the results of k^* and m^* , respectively, which show that k^* remains similar to $k_F = 8.1 \mu\text{m}^{-1}$ across the whole crossover. We note that the temperature of the cloud increases on the far BEC side, possibly due to heating introduced by the nearby confinement-induced p -wave resonance²⁹. We model our data on the BEC side of the resonance with a thermal ensemble of singlet pairs using the exact expression of the pair wavefunction in two dimensions and the experimentally determined temperatures³ (see Methods). The approximation of phase-disordered pairs in the normal state includes the correct short-range physics but neglects interactions between pairs as well as Pauli-blocking in the dissociation process. Our numerical data show that for experimentally

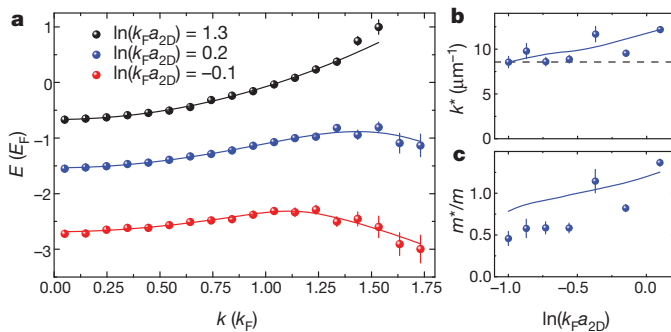


Figure 3 | Interaction dependence of the quasiparticle dispersion. **a**, Dispersion for different interaction strengths at $T/T_F = 0.27$. In the strongly interacting crossover regime, we observe back-bending of the dispersion relation, which vanishes in the regime of small interaction. The solid lines are the fits of a BCS-like dispersion relation, defined in the text. **b**, Wave vector k^* at which the back-bending of the spectral function occurs, extracted from a fit to the dispersion relation. The dashed line indicates the Fermi wave vector. **c**, Effective mass, extracted from a fit to the dispersion relation. The solid lines in **b** and **c** are the prediction of a thermal singlet model. The error bars (1σ) reflect the fitting errors.

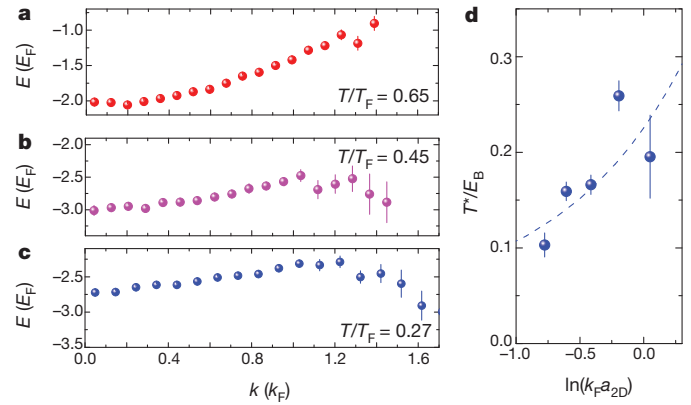


Figure 4 | Temperature dependence of the quasiparticle dispersion. **a–c**, Temperature dependence of the back-bending signature of the quasiparticle dispersion at $\ln(k_F a_{2D}) = -0.2$. **d**, Evolution of the pairing temperature T_B^* as a function of the interaction parameter $\ln(k_F a_{2D})$. The dashed line is the theoretical prediction of T^* from mean-field theory, scaled by a factor of 0.36.

relevant parameters the thermal singlet model can give rise to a back-bending feature of the photoemission spectrum with similar values of m^* and k^* as in the experiment (see solid lines in Fig. 3b and c). A contribution to the back-bending effect from incoherent pairs has been predicted previously in three dimensions³⁰. We note that, despite working reasonably well for k^* and m^* , the thermal singlet model makes incorrect predictions for the energy threshold in Fig. 2.

It is instructive to observe what happens to the dispersion signal at elevated temperatures. To this end, we have performed experiments in the range $0.27 \leq T/T_F \leq 0.65$ in the strongly interacting crossover regime. We observe that the BCS-like dispersion relation converts into a free-particle-like dispersion as the temperature is increased (see Fig. 4a–c) and we determine the temperature T_B^* at which this change takes place. Figure 4d shows the variation of T_B^* (in units of E_B) in the crossover regime and on the BEC side. We find that the temperature T_B^* is comparable to the pairing temperature T^* of mean-field theory (dashed line in Fig. 4d), but a factor of 0.36 smaller. This discrepancy could suggest that the appearance of the back-bending feature in the spectral function, which has been interpreted as a signature for many-body pairing^{3,11}, is mainly a qualitative evidence. In contrast, our spectroscopic measurements of the energy threshold (see Fig. 2) give quantitative results for many-body pairing and agreement with theory even in a regime where no back-bending feature is observable.

In conclusion, we have observed a many-body pairing gap above the superfluid transition temperature of a strongly interacting two-dimensional Fermi gas. Our results represent a major step in emulating and understanding pairing in two-dimensional, strongly correlated materials using cold atoms. In future experiments, an even closer cross-link to copper oxide high-temperature superconductors could be provided by the inclusion of an optical lattice potential to realize the two-dimensional Fermi–Hubbard model. In this model, a pseudogap phase is expected for a slightly doped antiferromagnetic state, and it is believed to precede exotic quantum states like the d -wave superconducting phase, which will be detectable by our technique.

METHODS SUMMARY

In our experimental set-up¹⁹, we prepare a quantum degenerate Fermi gas of ^{40}K atoms in a 50/50 mixture of the two lowest hyperfine states, $|F=9/2, m_F=-9/2\rangle$ and $|F=9/2, m_F=-7/2\rangle$. We confine the quantum gas to two dimensions in a deep optical lattice formed by a standing wave laser field, preparing approximately 30 layers. The interaction strength between spin-up and spin-down particles is tuned at a Feshbach resonance near 202.1 G. The photoemission measurement couples the $|F=9/2, m_F=-7/2\rangle$ state to the weakly interacting state $|F=9/2, m_F=-5/2\rangle$ using a radio-frequency photon of frequency Ω with negligible momentum transfer. We measure the momentum distribution of the transferred

atoms in a time-of-flight experiment and average the absorption signal azimuthally to obtain $A(k, \Omega)$, where $k = \sqrt{k_x^2 + k_y^2}$.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 September; accepted 11 October 2011.

- Loktev, V. M., Quick, R. M. & Sharapov, S. Phase fluctuations and pseudogap phenomena. *Phys. Rep.* **349**, 1–123 (2001).
- Dao, T.-L., Georges, A., Dalibard, J., Salomon, C. & Carusotto, I. Measuring the one-particle excitations of ultracold Fermionic atoms by stimulated Raman spectroscopy. *Phys. Rev. Lett.* **98**, 240402 (2007).
- Stewart, J. T., Gaebler, J. P. & Jin, D. S. Using photoemission spectroscopy to probe a strongly interacting Fermi gas. *Nature* **454**, 744–747 (2008).
- Damascelli, A., Hussain, Z. & Shen, Z.-X. Angle-resolved photoemission studies of the cuprate superconductors. *Rev. Mod. Phys.* **75**, 473–541 (2003).
- Ding, H. *et al.* Spectroscopic evidence for a pseudogap in the normal state of underdoped high- T_c superconductors. *Nature* **382**, 51–54 (1996).
- Randeria, M., Trivedi, N., Moreo, A. & Scalettar, R. T. Pairing and spin gap in the normal state of short coherence length superconductors. *Phys. Rev. Lett.* **69**, 2001 (1992).
- Trivedi, N. & Randeria, M. Deviations from Fermi liquid behavior above T_c in two dimensional short coherence length superconductors. *Phys. Rev. Lett.* **75**, 312 (1995).
- Bloch, I., Dalibard, J. & Zwirger, W. Many-body physics with ultracold gases. *Rev. Mod. Phys.* **80**, 885–964 (2008).
- Tsuchiya, S., Watanabe, R. & Ohashi, Y. Single-particle properties and pseudogap effects in the BCS-BEC crossover regime of an ultracold Fermi gas above T_c . *Phys. Rev. A* **80**, 033613 (2009).
- Hausmann, R., Punk, M. & Zwirger, W. Spectral functions and rf response of ultracold Fermionic atoms. *Phys. Rev. A* **80**, 063612 (2009).
- Gaebler, J. P. *et al.* Observation of pseudogap behaviour in a strongly interacting Fermi gas. *Nature Phys.* **6**, 569–573 (2010).
- Hu, H., Liu, X.-J., Drummond, P. D. & Dong, H. Pseudogap pairing in ultracold Fermi atoms. *Phys. Rev. Lett.* **104**, 240407 (2010).
- Chien, C.-C., Guo, H., He, Y. & Levin, K. Comparative study of BCS-BEC crossover theories above T_c : the nature of the pseudogap in ultracold atomic Fermi gases. *Phys. Rev. A* **81**, 023622 (2010).
- Nascimbène, S. *et al.* Fermi-liquid behavior of the normal phase of a strongly interacting gas of cold atoms. *Phys. Rev. Lett.* **106**, 215303 (2011).
- Sommer, A., Ku, M., Roati, G. & Zwierlein, M. W. Universal spin transport in a strongly interacting Fermi gas. *Nature* **472**, 201–204 (2011).
- Wulin, D., Guo, H., Chien, C.-C. & Levin, K. Spin transport in cold Fermi gases: A pseudogap interpretation of spin diffusion experiments at unitarity. *Phys. Rev. A* **83**, 061601 (2011).
- Pieri, P. *et al.* Pairing-gap, pseudogap, and no-gap phases in the radio-frequency spectra of a trapped unitary ^6Li gas. *Phys. Rev. A* **84**, 011608 (2011).
- Petrov, D. & Shlyapnikov, G. Interatomic collisions in a tightly confined Bose gas. *Phys. Rev. A* **64**, 012706 (2001).
- Fröhlich, B. *et al.* Radiofrequency spectroscopy of a strongly interacting two-dimensional Fermi gas. *Phys. Rev. Lett.* **106**, 105301 (2011).
- Randeria, M., Duan, J.-M. & Shieh, L.-Y. Bound states, Cooper pairing, and Bose condensation in two dimensions. *Phys. Rev. Lett.* **62**, 981–984 (1989).
- Miyake, K. Fermi liquid theory of dilute submonolayer ^3He on thin ^4He II film. *Prog. Theor. Phys.* **69**, 1794–1797 (1983).
- Drechsler, M. & Zwirger, W. Crossover from BCS-superconductivity to Bose-condensation. *Ann. Phys.* **504**, 15–23 (1992).
- Bertaina, G. & Giorgini, S. BCS-BEC crossover in a two-dimensional Fermi gas. *Phys. Rev. Lett.* **106**, 110403 (2011).
- Gorkov, L. & Melik-Barkhudarov, T. Contribution to the theory of superfluidity in an imperfect Fermi gas. *Sov. Phys. JETP* **13**, 1018–1022 (1961).
- Petrov, D. S., Baranov, M. A. & Shlyapnikov, G. V. Superfluid transition in quasi-two-dimensional Fermi gases. *Phys. Rev. A* **67**, 031601 (2003).
- Botelho, S. S. & Sá de Melo, C. A. R. Vortex-antivortex lattice in ultracold Fermionic gases. *Phys. Rev. Lett.* **96**, 040404 (2006).
- Shin, Y., Schunck, C. H., Schirotzek, A. & Ketterle, W. Tomographic rf spectroscopy of a trapped Fermi gas at unitarity. *Phys. Rev. Lett.* **99**, 090403 (2007).
- Regal, C., Greiner, M. & Jin, D. Observation of resonance condensation of Fermionic atom pairs. *Phys. Rev. Lett.* **92**, 040403 (2004).
- Günter, K., Stöferle, T., Moritz, H., Köhl, M. & Esslinger, T. *p*-wave interactions in low-dimensional Fermionic gases. *Phys. Rev. Lett.* **95**, 230401 (2005).
- Schneider, W. & Randeria, M. Universal short-distance structure of the single-particle spectral function of dilute Fermi gases. *Phys. Rev. A* **81**, 021601 (2010).

Acknowledgements We thank A. Georges, C. Kollath, D. Pertot, D. Petrov, M. Randeria, W. Zwirger and M. Zwierlein for discussions. The work was supported by EPSRC (EP/G029547/1), Daimler-Benz Foundation (B.F.), Studienstiftung and DAAD (M.F.).

Author Contributions The experimental set-up was devised and constructed by M.F., B.F., E.V. and M. Köhl, data-taking was performed by M.F., B.F., E.V. and M. Koschorreck, data analysis was performed by M.F., B.F. and M. Koschorreck, numerical modelling was performed by B.F., and the manuscript was written by M. Köhl with contributions from all co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M. Köhl (mk540@cam.ac.uk).

METHODS

Experimental set-up. We evaporatively cool a 50/50 spin mixture of ^{40}K atoms in the $|F = 9/2, m_F = -9/2\rangle \equiv |-9/2\rangle$ and $|F = 9/2, m_F = -7/2\rangle \equiv |-7/2\rangle$ states of the hyperfine ground state manifold¹⁷. After reaching quantum degeneracy in a crossed-beam optical dipole trap with approximately 70,000 atoms per spin state, we turn on an optical lattice potential in order to prepare two-dimensional Fermi gases^{19,29,31,32}. The optical lattice is formed by a horizontally propagating, retro-reflected laser beam of wavelength $\lambda = 1,064$ nm, focused to a waist of $140\ \mu\text{m}$. We increase the laser power over a time of 200 ms to reach a final potential depth of up to $V_{\text{lat}} = 83E_{\text{rec}}$, which is calibrated by intensity modulation spectroscopy. $E_{\text{rec}} = \hbar^2/(2m\lambda^2)$ is the recoil energy. The trapping frequency along the strongly confined direction is $\omega = 2\pi \times 78.5$ kHz. After loading the optical lattice, we adiabatically reduce the power of the optical dipole trap such that the atoms are confined only by the Gaussian intensity envelope of the lattice laser beams. The radial trapping frequency of the two-dimensional gases is $\omega_{\perp} = 2\pi \times 127$ Hz for $V_{\text{lat}} = 83E_{\text{rec}}$ and we confine of the order of 10^3 atoms per two-dimensional gas at the centre of the trap. Along the axial direction we populate approximately 30 layers of the optical lattice potential with an inhomogeneous peak density distribution. Approximately two thirds of the two-dimensional layers with highest density dominate the measured signal, and their relevant energy scales E_F , E_B and $\Delta^2/2E_F$ are more than an order of magnitude larger than the trapping frequency ω_{\perp} . Therefore, finite particle number effects do not influence the measured signal. After evaporation, we adiabatically increase the interaction strength by lowering the magnetic field, at a rate of up to $0.25\ \text{G ms}^{-1}$, to a value near the Feshbach resonance at 202.1 G. We apply a radio-frequency pulse near 47 MHz with a Gaussian amplitude envelope with a full-width at half-maximum of $230\ \mu\text{s}$ to transfer atoms from the $|-7/2\rangle$ state to the $|F = 9/2, m_F = -5/2\rangle$ state. Atoms in the $|9/2, -5/2\rangle$ state have a two-body s -wave scattering length of 130 Bohr radii with the $|-7/2\rangle$ state and 250 Bohr radii with the $|-9/2\rangle$ state³. We turn off the optical lattice 100 μs after the radio-frequency pulse, switch off the magnetic field and apply a magnetic field gradient to achieve spatial splitting of the three spin components in a Stern-Gerlach experiment. For each run, the magnetic field is calibrated using spin-rotation with an r.f. pulse of an imbalanced mixture on the $|-9/2\rangle/|-7/2\rangle$ transition. The magnetic field accuracy deduced from these measurements is <3 mG. We measure the temperature by ballistic expansion of a weakly interacting gas, and the quoted numbers refer to the average of T/T_F across the whole sample.

Determination of the energy threshold E_{th} of the energy distribution curve. We fit our data with a double-peak fitting function comprising a Gaussian for the

atomic signal and a modified Gumbel function $f(\Omega) = \alpha \exp[-(\Omega - \Omega_0)/b - a \exp(-(\Omega - \Omega_0)/(ab))]$ for the pairing peak. The parameter Ω_0 measures the peak position and the parameters a and b measure skewness and width. For our further analysis, we only use the peak position Ω_0 , which does not depend on the line shape function used. From this fit we determine the maximum of the molecular peak $\nu_{\text{max}} = \Omega_0$ and the minimum between the atomic and the molecular peak ν_{min} . Between $\nu_1 = \nu_{\text{max}}$ and $\nu_2 = \nu_{\text{min}} - 2$ kHz we fit the data with a linear function and determine the zero-crossing of the linear extrapolation as the energy threshold E_{th} . We correct the obtained result for our spectral resolution of 1.5 kHz, obtained from the width of the Gaussian fits. The data are normalized to the two-body binding energy in vacuum which we obtain from the transcendental equation⁸

$$l_z/a = \int_0^\infty \frac{du}{\sqrt{4\pi u^3}} \left(1 - \frac{\exp(-E_B u/(\hbar\omega))}{\sqrt{(1 - \exp(-2u))/2u)}} \right) \quad (1)$$

Here, $l_z = \sqrt{\hbar/m\omega}$ and a is the three-dimensional scattering length using the following parameters of the Feshbach resonance: $B_0 = 202.1$ G, $\Delta B = 7$ G and $a_{\text{BG}} = 174a_B$ where a_B is the Bohr radius.

Thermal singlet model. We model our data on the BEC side of the resonance with a thermal ensemble of singlet pairs³. The expression for the wavefunction of the bound state in two dimensions is $\psi_B(r) = \sqrt{2/a_{2D}} K_0(r/a_{2D})$, in which $K_0(x)$ is the modified Bessel function, and for the scattering state is $\psi_q(r) = J_0(qr) - \frac{if(q)}{4} H_0^{(1)}(qr)$, in which $J_0(x)$ is the Bessel function of the first kind and $H_0^{(1)}(x)$ is the Hankel function of the first kind¹⁸. $f(q)$ is the scattering amplitude between the state $|-7/2\rangle$ and the final state $|-5/2\rangle$. We compute the momentum resolved r.f. spectrum for the dissociation from the bound state to the scattering state, averaging over a thermal distribution of the centre-of-mass momenta of the initial pairs using Monte Carlo sampling. From the momentum-resolved r.f. spectrum we calculate the effective mass m^* and the wave vector k^* using the same fitting routines as for the experimental data. This model of tightly bound pairs in the normal state includes the correct short-range physics but neglects many-body pairing, interactions between atoms and between pairs, as well as quantum statistical effects. Therefore, we do not expect quantitative agreement in the strongly interacting regime or on the BCS side of the resonance.

31. Martiyanov, K., Makhalov, V. & Turlapov, A. Observation of a two-dimensional Fermi gas of atoms. *Phys. Rev. Lett.* **105**, 030404 (2010).
32. Dyke, P. *et al.* Crossover from 2d to 3d in a weakly interacting Fermi gas. *Phys. Rev. Lett.* **106**, 105304 (2011).

The oxidation state of Hadean magmas and implications for early Earth's atmosphere

Dustin Trail^{1,2}, E. Bruce Watson^{1,2} & Nicholas D. Tailby^{1,2}

Magmatic outgassing of volatiles from Earth's interior probably played a critical part in determining the composition of the earliest atmosphere, more than 4,000 million years (Myr) ago¹. Given an elemental inventory of hydrogen, carbon, nitrogen, oxygen and sulphur, the identity of molecular species in gaseous volcanic emanations depends critically on the pressure (fugacity) of oxygen. Reduced melts having oxygen fugacities close to that defined by the iron–wüstite buffer would yield volatile species such as CH₄, H₂, H₂S, NH₃ and CO, whereas melts close to the fayalite–magnetite–quartz buffer would be similar to present-day conditions and would be dominated by H₂O, CO₂, SO₂ and N₂ (refs 1–4). Direct constraints on the oxidation state of terrestrial magmas before 3,850 Myr before present (that is, the Hadean eon) are tenuous because the rock record is sparse or absent. Samples from this earliest period of Earth's history are limited to igneous detrital zircons that pre-date the known rock record, with ages approaching ~4,400 Myr (refs 5–8). Here we report a redox-sensitive calibration to determine the oxidation state of Hadean magmatic melts that is based on the incorporation of cerium into zircon crystals. We find that the melts have average oxygen fugacities that are consistent with an oxidation state defined by the fayalite–magnetite–quartz buffer, similar to present-day conditions. Moreover, selected Hadean zircons (having chemical characteristics consistent with crystallization specifically from mantle-derived melts) suggest oxygen fugacities similar to those of Archaean and present-day mantle-derived lavas^{2–4,9,10} as early as ~4,350 Myr before present. These results suggest that outgassing of Earth's interior later than ~200 Myr into the history of Solar System formation would not have resulted in a reducing atmosphere.

The oxygen fugacity (f_{O_2}) of a magma influences mineral saturation, element partitioning and the molecular speciation of gases exsolved on eruption^{1–4}. Traditionally, the oxidation state of ancient eruptives is assessed by measuring Fe³⁺/Fe²⁺ values of unaltered glass samples¹¹. For less pristine ancient samples, researchers rely on changes in crystal–liquid partitioning versus f_{O_2} for heterovalent transition elements such as V and Cr that are generally regarded as immobile during rock alteration^{2,3}. Studies of this nature have provided consistent lines of evidence that the upper mantle reached its present-day oxidation state by 3.6 Gyr ago and perhaps as early as 3.9 Gyr (ref. 3). Whereas the final oxidation state of a glass or mineral reflects that of the parent magma, subsequent metamorphism or alteration at Earth's surface may reset the primary redox equilibria. Preservation of the redox state of the most primitive Hadean magmas therefore requires a physically and chemically robust repository. Fortunately, zircon is an exceptionally durable mineral that retains primary chemistry for most elements and isotopes from the time of igneous crystallization¹², and zircon crystals are the only known terrestrial solids dating from the first 500 Myr of Earth history. Zircon thus presents a unique opportunity to evaluate the oxidation state of terrestrial melts before the Archaean if the partitioning of a variable-valence cation into zircon can be characterized as a function of f_{O_2} .

Here we report a calibration to determine the oxidation state of a magmatic melt based on the incorporation of cerium (Ce) into zircon. Cerium is unique among elements that partition into zircon (and among the rare-earth elements, REEs) in that it can exist in melts as either Ce⁴⁺ or Ce³⁺. The magnitude of Ce enrichment in zircon depends on the Ce⁴⁺/Ce³⁺ ratio of the medium from which it crystallizes. Because Ce⁴⁺ is vastly more compatible than Ce³⁺ in zircon, more-oxidized melts will yield higher Ce contents in the mineral. To explore the details of Ce uptake, zircons were grown in a piston cylinder apparatus at a pressure of 1 GPa and at temperatures of 900–1,300 °C from hydrous silicate melts (~72 wt% SiO₂) doped with lanthanum (La), Ce and praseodymium (Pr) (with and without phosphorus). Lanthanum and Pr were included as 'bracketing' elements because these exist only as 3+ ions, and partition coefficients for zircon trivalent REEs are known to monotonically increase from La to lutetium (Lu) (ref. 13); by convention, partition coefficients are expressed as the crystal/melt concentration ratio, $D_{REE}^{zrc/melt}$ (here zrc indicates zircon, the crystal concerned). This systematic behaviour provides a convenient reference for natural zircon to characterize the relative enrichment of Ce in zircon, relative to other REEs, with changes in f_{O_2} . This is expressed as a 'Ce anomaly', $(Ce/Ce^*)_D$, and is calculated in the following manner:

$$\left(\frac{Ce}{Ce^*}\right)_D = \frac{D_{Ce}^{zrc/melt}}{\sqrt{D_{La}^{zrc/melt} \times D_{Pr}^{zrc/melt}}} \quad (1)$$

where $D_{Ce}^{zrc/melt}$ is the combined partition coefficient of Ce⁴⁺ and Ce³⁺ in zircon, and $D_{La}^{zrc/melt}$ and $D_{Pr}^{zrc/melt}$ are respectively the partition coefficients for trivalent La and Pr. The denominator on the right hand side (equal to Ce*) is the interpolated value for the Ce³⁺ partition coefficient based on the partition coefficients of bracketing elements La and Pr. Accordingly, if $(Ce/Ce^*)_D \approx 1$, then Ce⁴⁺ must have been effectively absent in the melt at the time of zircon crystallization.

The Ce⁴⁺/Ce³⁺ ratio of the melt was varied by imposing different oxygen fugacities on the experimental charges, broadly corresponding to the range observed in terrestrial magmas. The change in element partitioning in response to differences in f_{O_2} was then measured in synthesized zircon populations and melts using an electron microprobe. The results show that Ce anomalies are inversely related to temperature (Fig. 1) and can be expressed by the following empirical equation:

$$\ln\left(\frac{Ce}{Ce^*}\right)_D = (0.1156 \pm 0.0050) \times \ln(f_{O_2}) + \frac{13,860 \pm 708}{T} - 6.125 \pm 0.484 \quad (2)$$

where f_{O_2} is the oxygen fugacity and T is temperature in K (T can be estimated for natural samples using the Ti-in-zircon thermometer⁷).

Phosphorus substitution for silicon is known to compensate the charge imbalance accompanying the entry of trivalent REEs into

¹Department of Earth and Environmental Sciences, Rensselaer Polytechnic Institute, Troy, New York 12180, USA. ²New York Center for Astrobiology, Rensselaer Polytechnic Institute, Troy, New York 12180, USA.

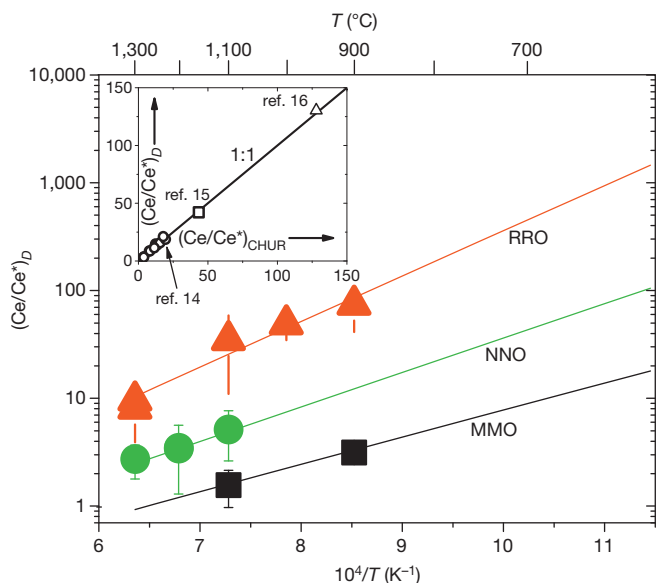


Figure 1 | Experimental calibration, relating melt oxygen fugacity to zircon Ce anomalies and temperature. On average, six zircons and five glasses were analysed for each of the 11 experiments. (Two experiments each at 1,300 °C/RRO and 1,200 °C/NNO yield virtually identical Ce anomalies and plot on top of each other.) The Ru–RuO₂ (RRO) buffer defines an oxygen fugacity (f_{O_2}) that is ~ 1 log unit greater than that of the magnetite–haematite (MH) buffer at $\sim 1,000$ °C, Ni–NiO (NNO) is ~ 0.8 log units greater than FMQ, and the Mo–MoO₂ (MMO) buffer is nearly identical to IW (<0.5 log units) at all temperatures. The curves drawn are fits to the data according to equation (2). Inset, Ce anomalies from natural zircon–melt partitioning studies^{14–16} (reference numbers are shown); calculated partition coefficients (subscript ‘D’) are compared to those normalized to chondrite (subscript ‘CHUR’). Error bars, 1σ .

zircon, but the addition of 0.15–0.8 wt% P₂O₅ to three experiments did not lead to differences in $(Ce/Ce^*)_D$. In addition, two experiments resulted in the saturation of light-REE-bearing phases allanite and monazite, but showed no detectable change in $(Ce/Ce^*)_D$ of zircon. These observations suggest that the calibration is unaffected by the presence of light-REE phases or the P content of the melt.

Before applying equation (2) to natural samples, it is important to note that in most natural settings $(Ce/Ce^*)_D$ cannot be calculated because the REE concentration of the melt at the time of zircon saturation is unknown. As an alternative, Ce anomalies are often calculated by normalizing to the chondritic uniform reservoir (CHUR), in which case the relationship between $(Ce/Ce^*)_D$ and $(Ce/Ce^*)_{CHUR}$ must be established. Plotting $(Ce/Ce^*)_D$ versus $(Ce/Ce^*)_{CHUR}$ for natural zircon–melt partitioning studies^{14–16} demonstrates a near perfect 1:1 correlation (Fig. 1 inset), suggesting that $(Ce/Ce^*)_D \approx (Ce/Ce^*)_{CHUR}$ is a reasonable assumption for out-of-context zircons. In other words, the magnitude of the Ce anomaly will be independent of the concentration of trace elements in the melt as long as the light REEs are not fractionated relative to one another.

If the assumptions outlined above are valid, then zircon data^{6,17} give an average f_{O_2} of 0.5 log units below the value defined by the fayalite–magnetite–quartz (FMQ) buffer (± 2.3 log units) for Hadean melts, excluding the single outlier that plots well below the iron–wüstite (IW) buffer curve (Fig. 2). Lunar zircon source melts yield a calculated oxygen fugacity at IW (± 1.3 log units), which is within error of previous estimates from lunar rocks and minerals¹⁸. Also note that lunar zircon Ce anomalies show a clear inverse relationship with temperature, in agreement with the prediction of equation (2). There is no correlation between temperature or age versus difference from the IW buffer for lunar zircons, though the Hadean zircons exhibit a weak trend towards more reducing conditions with lower crystallization temperatures.

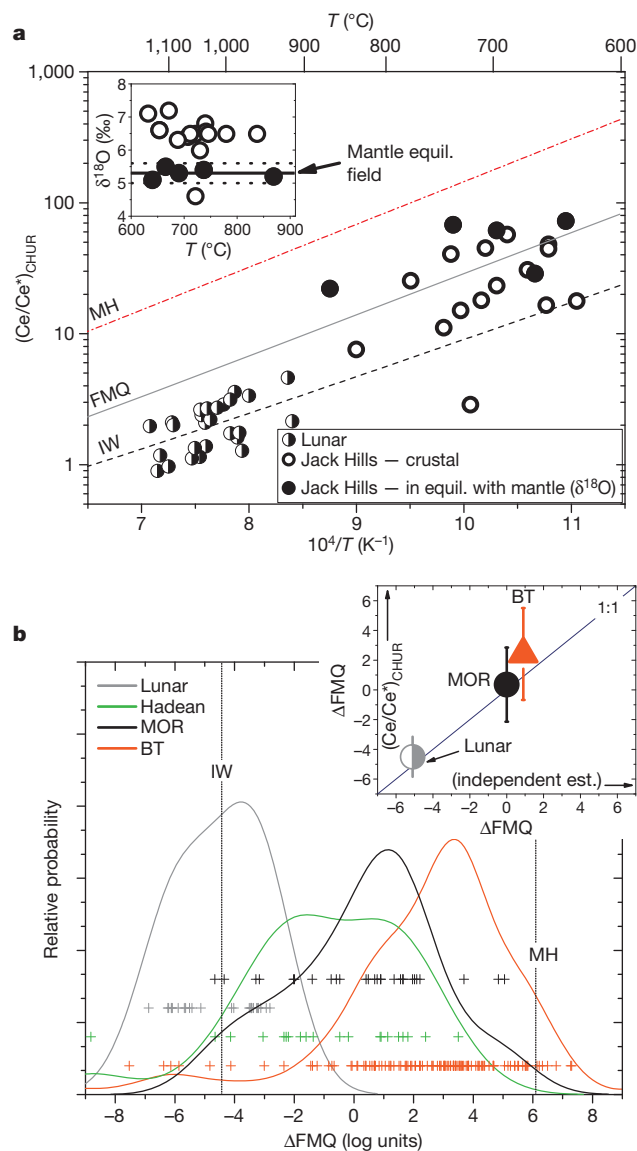


Figure 2 | Oxygen fugacity and oxygen isotope plots for natural samples using the experimental calibration. a, Main panel; $(Ce/Ce^*)_{CHUR}$ versus $1/T$ for Jack Hills Hadean and lunar zircons with MH, FMQ and IW buffer curves (equation (2)). Inset, $\delta^{18}O$ versus temperature. Filled symbols, crystals that have $\delta^{18}O$ values in the mantle equilibrium field for zircon ($n = 5$)²⁰ (inset), samples shown by open symbols yield $\delta^{18}O$ values more akin to a crustal origin ($n = 15$). The open and filled symbols in the main panel and the inset are from the same samples. b, Oxygen fugacity probability distribution plots for lunar ($n = 27$)¹⁹, Hadean ($n = 20$)^{6,17}, mantle-derived mid-ocean ridge (MOR) residual liquids ($n = 27$)^{22–24} and Bishop Tuff (BT) magma-chamber ($n = 120$)³⁰ zircons. Values of f_{O_2} are given as difference from FMQ in log units (ΔFMQ). The coloured symbols correspond to calculated f_{O_2} values from individual zircons. Zircon host melts have calculated f_{O_2} values for the aforementioned samples of IW (± 1.3), FMQ -0.5 (± 2.3), FMQ $+0.4$ (± 2.6), FMQ $+2.4$ (± 3.1), respectively. The calculated f_{O_2} values are in agreement with independent f_{O_2} estimates for BT³⁰, MOR¹⁰ and the Moon¹⁸ (inset). The errors are in log units and they represent the standard deviations of f_{O_2} values from individual sample sets.

Accepting for the moment that terrestrial Hadean zircons probably crystallized under relatively oxidizing conditions (when compared to lunar zircons), the remaining question is whether they truly reflect the prevailing f_{O_2} of Earth’s early mantle, which was the dominant source of volcanic emanations forming the early atmosphere. Lunar zircons did in fact crystallize from mantle-derived melts—specifically the residual liquid fraction of the lunar magma ocean¹⁹. Most terrestrial Hadean zircons, on the other hand, crystallized from crust-derived

felsic melts, as indicated by multiple geochemical signatures^{6,8,9}. Oxygen isotopes, for example, demonstrate that the vast majority of terrestrial Hadean zircons fall outside the zircon mantle equilibrium field ($\delta^{18}\text{O}_{\text{VSMOW}} = 5.3 \pm 0.3\text{‰}$ (1σ)^{5,20}, which is a strong argument for the input of supracrustal material into the source melts. Moreover, most zircon crystallization temperatures are consistent with formation in water-saturated or near-water-saturated melts^{7,8}.

Conclusions about the oxidation state of the Hadean mantle based on zircons of crustal affinity are admittedly indirect. However, a subset of the Hadean zircon population does appear to be mantle-derived. (We note that direct partial melts of the mantle cannot crystallize zircon, but the fractionation products (residuals) of such melts do eventually saturate in zircon²¹.) Five of the Hadean crystals have $\delta^{18}\text{O}$ values that fall within the mantle equilibrium field, which has been shown to be constant ($\pm 0.2\text{‰}$) for the past 4.4 Gyr (ref. 20). If these $\delta^{18}\text{O}$ values are primary, then these zircons crystallized from uncontaminated, mantle-derived melts that did not interact with the hydrosphere. The five zircons in the $\delta^{18}\text{O}$ mantle equilibrium field—including two crystals with U–Pb ages approaching 4,400 Myr—give a calculated oxygen fugacity of FMQ + 1.4 (± 2) (Fig. 3). As a comparison, zircons crystallized from residual liquids of present-day mantle-derived mid-ocean ridge basalts^{22–24} yield calculated f_{O_2} values of FMQ + 0.4 (± 2.6), which agrees with estimates of the oxidation state of the upper mantle¹⁰. This also implies that the redox conditions in residual, zircon-saturated liquids of present-day mid-ocean ridge systems did not undergo significant changes in f_{O_2} (relative to the buffer curve) along the liquid line of descent from basalt. This is also consistent with observations from lunar zircons. Finally, we note that Martian basalts (for example, shergottite meteorite Dar al Gani 476) exhibiting little or no interaction with a crustal component are only about 1 log unit higher than the estimated f_{O_2} of the upper mantle²⁵. Furthermore, calculated oxygen fugacities from the same meteorite are all within a log unit of the average (FMQ – 2.5), even though phases cover a crystallization range of $>300\text{ °C}$.

As in all studies in which mantle f_{O_2} is estimated from mantle-derived glasses or minerals^{3,9}, it is important to emphasize that the calculated f_{O_2} may not directly reflect that of the mantle source region. We note, however, that mantle-derived Hadean zircons would have had to crystallize from residual melts that underwent ~ 4 log units of change in f_{O_2} to be reconcilable with mantle source regions in equilibrium with the IW buffer. At present and in the Archaean, relative f_{O_2}

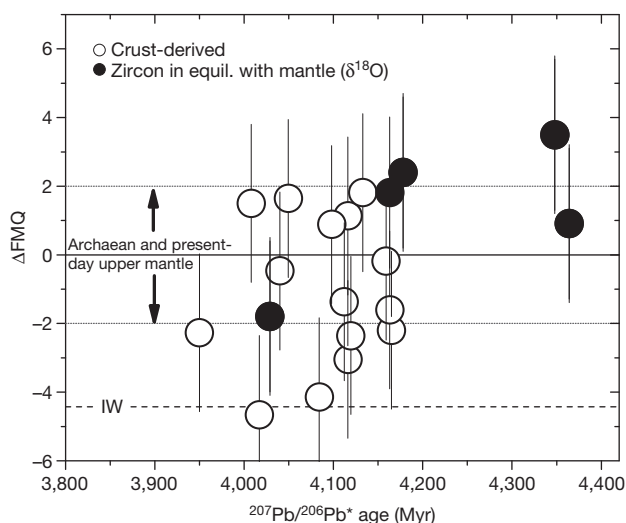


Figure 3 | Oxygen fugacities of Hadean melts plotted against zircon crystallization age. Errors for individual points are based on the standard deviation of the entire data set ($n = 20$; $1\sigma = \pm 2.3$ log units). Zircons with mantle signatures are within error of the estimated oxidation state of the present-day and Archaean upper mantle^{3,10,26}. On average, oxygen fugacities are lower for younger zircons, though this trend is not robust given the present data set.

changes along magmatic liquid lines of descent are broadly of the order of ~ 1 log unit^{4,10,26,27}. If the evolution of Hadean melts was similarly constrained, then our results imply that the mantle reached its present-day oxidation state $\sim 4,350$ Myr ago (Fig. 3). This interpretation is consistent with core formation models calling for physical and chemical isolation from the (upper) mantle ~ 30 Myr after planet formation²⁸.

If our deductions regarding the oxidation state of Hadean magmas are correct, then the speciation of gases emanating from the Earth at this time would have been dominated by CO_2 , SO_2 , H_2O and N_2 (ref. 26). An atmosphere of this composition is known to yield a lower abundance of sugars and especially amino acids and nucleotides¹. If a highly reduced atmosphere is required for the origin of life, then it may have occurred exceptionally early on our planet. However, pre-4,400-Myr outgassing of H_2 coupled with slow escape²⁹ may have resulted in an atmosphere out of equilibrium with Earth's interior. Alternatively, a 'late veneer' may have served as a source of pre-biotic molecules. These alternative scenarios notwithstanding, our results offer the first glimpse into the oxidation state of terrestrial magmas before 3,900 Myr ago, and the data presented here open up the possibility of exploring the oxidation state of zircon-bearing rocks throughout the geologic record. This calibration will be especially valuable in the early Archaean or Hadean eons, for which zircon is often the only source of primary chemical information.

METHODS SUMMARY

Starting materials were contained in platinum (Pt) or silver (Ag) capsules, and f_{O_2} was buffered by placing ruthenium and ruthenium dioxide ($\text{Ru} + \text{RuO}_2$; RRO) in direct physical contact with the experimental charge or by use of a separate buffer chamber separated by a Pt or silver-palladium (Ag–Pd) membrane (NNO or MMO). Quenched glasses were analysed by electron microprobe, with at least one measurement made at the top and bottom of the capsule to assess melt homogeneity. Synthesized zircons were small ($\sim 3\text{--}20\text{ }\mu\text{m}$) and in some cases the glass was dissolved with hydrofluoric acid (HF) and the zircons recovered in order to obtain more accurate measurements.

For the inset plot of Fig. 1, $\text{Pr}_x = 2/3(\text{La})_x + 1/3(\text{Nd})_x$ for ref. 14, where x represents a normalized value, either to CHUR or the melt. Ce anomalies for natural zircons^{6,19,23–25,30} are calculated in a similar fashion to equation (1) in the text: $(\text{Ce}/\text{Ce}^*)_{\text{CHUR}} = \text{Ce}_{\text{CHUR}}/(\text{La}_{\text{CHUR}} \times \text{Pr}_{\text{CHUR}})^{1/2}$, where La_{CHUR} and so on represents normalization to the chondritic uniform reservoir rather than the melt concentration. Ce anomalies are plotted for Hadean zircons with U–Pb ages that are $\geq 94\%$ concordant, and with REE+Ti measurements that reflect primary igneous chemistry^{5,6,17}. Zircon crystallization temperatures in Fig. 2 are calculated assuming unity Ti-activity⁷; sub-unity Ti activity (for example, ~ 0.5) results in systematic shifts of all data to higher values of f_{O_2} , though shifts are within error of the standard deviation of each population. Lunar zircons are from Apollo 14 polymict breccias 14304, 14305 and 14321¹⁹ and are normalized to an average f_{O_2} of IW at 700 °C , for direct comparison with the other data in the probability distribution plot of Fig. 2. For the inset of Fig. 2b, average f_{O_2} values calculated from zircon Ce anomalies are plotted against independent estimates from lunar basaltic igneous rocks (IW – 0.6)¹⁸, terrestrial mid-ocean ridge basalts¹⁰ (FMQ), and average oxygen fugacities returned from Bishop Tuff titanomagnetite-ilmenite pairs (NNO)²⁷.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 22 March; accepted 6 October 2011.

- Kasting, J. F. Earth's early atmosphere. *Science* **259**, 920–926 (1993).
- Canil, D. Vanadium partitioning and the oxidation state of Archaean komatiite magmas. *Nature* **389**, 842–845 (1997).
- Delano, J. W. Redox history of the Earth's interior since ~ 3900 Ma: implications for prebiotic molecules. *Orig. Life Evol. Biosph.* **31**, 311–341 (2001).
- Burgisser, A. & Scaillet, B. Redox evolution of a degassing magma rising to the surface. *Nature* **445**, 194–197 (2007).
- Cavosie, A. J., Valley, J. W., Wilde, S. A., & Edinburgh Ion Microprobe Facility. Magmatic $\delta^{18}\text{O}$ in 4400–3900 Ma detrital zircons: a record of the alteration and recycling of crust in the early Archaean. *Earth Planet. Sci. Lett.* **235**, 663–681 (2005).
- Cavosie, A. J., Valley, J. W., Wilde, S. A., & Edinburgh Ion Microprobe Facility. Correlated microanalysis of zircon: trace element, $\delta^{18}\text{O}$, and U–Th–Pb isotopic constraints on the igneous origin of complex 3900 Ma detrital grains. *Geochim. Cosmochim. Acta* **69**, 637–648 (2006).

7. Watson, E. B. & Harrison, T. M. New thermometer reveals minimum melting conditions on earliest Earth. *Science* **308**, 841–844 (2005).
8. Hopkins, M., Harrison, T. M. & Manning, C. M. Low heat flow inferred from >4 Gyr zircons suggests Hadean plate boundary interactions. *Nature* **456**, 493–496 (2008).
9. Berry, A. J., Danyushevsky, L. V., O'Neill, H. St. C., Newville, M. & Sutton, S. R. Oxidation state of iron in komatiitic melt inclusions indicates hot Archaean mantle. *Nature* **455**, 960–963 (2008).
10. Mallmann, G. & O'Neill, H. St. C. The crystal/melt partitioning of V during mantle melting as a function of oxygen fugacity compared with some other elements (Al, P, Ca, Sc, Ti, Cr, Fe, Ga, Y, Zr and Nb). *J. Petrol.* **50**, 1765–1794 (2009).
11. Carmichael, I. S. E. & Ghiorso, M. S. The effect of oxygen fugacity on the redox state of natural liquids and their crystallizing phases. *Rev. Mineral.* **24**, 191–212 (1990).
12. Cherniak, D. J., Hancher, J. M. & Watson, E. B. Rare-earth diffusion in zircon. *Chem. Geol.* **134**, 289–301 (1997).
13. Rubatto, D. & Hermann, J. Experimental zircon/melt and zircon/garnet trace element partitioning and implications for the geochronology of crustal rocks. *Chem. Geol.* **241**, 38–61 (2007).
14. Thomas, J. B., Bodnar, R. J., Shimizu, N. & Sinha, A. K. Determination of zircon/melt trace element partition coefficients from SIMS analysis of melt inclusions in zircon. *Geochim. Cosmochim. Acta* **66**, 2887–2901 (2002).
15. Hinton, R. W. & Upton, B. G. J. The chemistry of zircon: variations within and between large crystals from syenite and alkali basalt xenoliths. *Geochim. Cosmochim. Acta* **55**, 3287–3302 (1991).
16. Sano, Y., Terada, K. & Fukuoka, T. High mass resolution ion microprobe analysis of rare earth elements in silicate glass, apatite and zircon: lack of matrix dependency. *Chem. Geol.* **184**, 217–230 (2002).
17. Fu, B. *et al.* Ti-in-zircon thermometry: applications and limitations. *Contrib. Mineral. Petrol.* **156**, 197–215 (2008).
18. Sato, M., Hickling, N. L. & McLane, J. E. in *Proc. Fourth Lunar Science Conference* Vol. 1 (ed. Grose, W. A.) 1061–1079 (Geochim Cosmochim Acta Suppl. 4, Pergamon, 1973).
19. Taylor, D. J., McKeegan, K. D. & Harrison, T. M. Lu–Hf zircon evidence for rapid lunar differentiation. *Earth Planet. Sci. Lett.* **279**, 157–164 (2009).
20. Valley, J. W. *et al.* 4.4 billion years of crustal maturation: oxygen isotope ratios of magmatic zircon. *Contrib. Mineral. Petrol.* **150**, 561–580 (2005).
21. Watson, E. B. & Harrison, T. M. Zircon saturation revisited: temperature and composition effects in a variety of crustal magma types. *Earth Planet. Sci. Lett.* **64**, 295–304 (1983).
22. Coogan, L. A. & Hinton, R. W. Do trace element compositions of detrital zircons require Hadean continental crust? *Geology* **34**, 633–636 (2006).
23. Schmitt, A. K. *et al.* Rapid cooling rates at an active mid-ocean ridge from zircon thermochronology. *Earth Planet. Sci. Lett.* **302**, 349–358 (2011).
24. Cavosie, A. J., Kita, N. K. & Valley, J. W. Primitive oxygen-isotope ratio recorded in magmatic zircon from the Mid-Atlantic Ridge. *Am. Mineral.* **94**, 926–934 (2009).
25. Herd, C. D. K., Borg, L. E. & Jones, J. H. Oxygen fugacity and geochemical variations in the Martian basalts: implications for Martian basalt petrogenesis and the oxidation state of the upper mantle of Mars. *Geochim. Cosmochim. Acta* **66**, 2025–2036 (2002).
26. Frost, D. J. & McCammon, C. A. The redox state of Earth's mantle. *Annu. Rev. Earth Planet. Sci.* **36**, 389–420 (2008).
27. Hildreth, W. & Wilson, C. J. N. Compositional zoning of the Bishop Tuff. *J. Petrol.* **48**, 951–999 (2007).
28. Kleine, T., Munker, C., Mezger, K. & Palme, H. Rapid accretion and early core formation on asteroids and the terrestrial planets from Hf–W chronometry. *Nature* **418**, 952–955 (2002).
29. Tian, F., Toon, O. B., Pavlov, A. A. & De Stereck, H. A hydrogen-rich early Earth atmosphere. *Science* **308**, 1014–1017 (2005).
30. Reid, M. R., Vazquez, J. A. & Schmitt, A. K. Zircon-scale insights into the history of a Supervolcano, Bishop Tuff, Long Valley, California, with implications for the Ti-in-zircon geothermometer. *Contrib. Mineral. Petrol.* **161**, 293–311 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by the NASA Astrobiology Institute (grant no. NNA09DA80A to The New York Center for Astrobiology).

Author Contributions E.B.W. identified the importance of investigating redox sensitive elements in zircon. D.T. designed the experiments and took measurements. D.T. wrote the manuscript and interpreted the data with significant contributions from E.B.W. and N.D.T.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to D.T. (traild@rpi.edu).

METHODS

Zircons were synthesized in silicate melts with water added as silicic acid, gibbsite, $\text{Zr}(\text{OH})_4 \pm \text{H}_2\text{O}$ (ref. 31). Experiments run from 1,000 to 1,300 °C contained ~8–10 wt% H_2O and those run at 900 °C contained 12 wt% H_2O , which was found to aid zircon synthesis at lower temperatures. Saturation of other REE-bearing phases was generally avoided by doping melt compositions with progressively lower REE concentrations for lower-temperature experiments. (Starting compositions can be found in Supplementary Table 1.)

The pressure cell used here is similar to that described in ref. 32. Capsules made of either Ag (900 °C) or Pt (1,000–1,300 °C) were pressure sealed with Pt or Ag–Pd gaskets respectively to ensure a water tight seal. The highest oxygen fugacity imposed on the experimental charges was comprised of a 9:1 mixture of RuO_2 :Ru added with the starting silicate mix into the capsule. This is a common buffering technique because RuO_2 is virtually insoluble in silicate melts. At run termination, the buffer was inspected to ensure the presence of RuO_2 and Ru, and that both phases were in clear chemical communication with the glass and synthesized zircons. Experiments buffered at NNO or MMO contained a chamber external to the experiment that contained the metal, oxide and H_2O . For experiments run at 1,000–1,300 °C, Pt capsules were surrounded by either a Mo or an oxidized Ni holder, while those at 900 °C were contained within a thick-walled Ag capsule, and the buffer was separated by a thin sheet of 30Pd–70Ag metal. These experiments are buffered by the permeability of H_2 through Pt or 30Pd–70Ag; that is, fixing the hydrogen fugacity (f_{H_2}) at MMO or NNO buffers the f_{O_2} in a similar manner to that described in ref. 33. A diagram of the experimental set-ups can be found in Supplementary Fig. 1.

Synthesized zircons (~3–20 µm) were analysed using the Rensselaer Polytechnic Institute Cameca SX100 electron microprobe. All elements were calibrated on standards for every electron microprobe session (accelerating voltage, 15 kV). The only interference was the slight overlap of the La L_β peak with the L_α line of Pr. This was removed by synthesizing a melt with 0–2 wt% of La_2O_3 so that the contribution of La L_β to Pr L_α counts could then be subtracted out. The major elements of the melts were analysed with a beam current of 10 nA and a 40–60 µm spot size, and trace elements with a 50 nA, 30 µm spot size (analysed glass compositions can be found in Supplementary Table 2).

Zircons were analysed with a focused beam and depending on crystal size, electron microprobe currents ranged from 20 to 200 nA; synthetic REE-phosphates and zircon were used as standards. Aluminium was used as a monitor for secondary fluorescence and in some cases after the melt was analysed, it was dissolved in HF, so that the recovered zircon fraction could be analysed. Single acquisition times as long as 1,000 s were used for the high-temperature experiments especially, which yielded the lowest partition coefficients. For some zircons,

multiple acquisitions were made on the same crystal to accumulate additional counts. In order to increase the counting time for the La L_α peak, essentially 3 of the 5 spectrometers (the counting time on one spectrometer was split between La and Ce) were devoted to simultaneous peak counting (large pentaerythritol (LPET) crystal), and accumulated totals were used to obtain a concentration. Another spectrometer was used for Si K_α , Zr L_α , Al K_α (thallium acid phthalate (TAP)) while the final was used for Pr (large lithium fluoride (LLIF)). Under this protocol, the detection limits for La, Ce and Pr are ~30, ~100 and ~80 p.p.m., respectively. Because of the very low counts on some unknowns, zircons were synthesized without REEs and measured under the same protocols to test for trace interferences on the peak and background positions. Zircon concentrations along with individual grains in glass and epoxy matrix are shown for comparison in Supplementary Table 3. The partition coefficients used to construct Fig. 2 of the main text are provided in Supplementary Table 4.

All data used to create Figs 2 and 3 are included in Supplementary Table 5. Two criteria were considered when applying our calibration to Hadean grains. First, it is known that Hadean zircons may record multiple episodes of growth over their >4 Gyr history, meaning that grains may be heterogeneous in age and chemistry. Our calibration is applied to zircons for which the following determinations were performed on the same location on the crystal: (1) U/Pb age, (2) oxygen isotope, (3) magmatic REE concentrations and (4) temperature (Ti-thermometry) data. Correlated geochemical data for the first three items can be found in ref. 6. These data are then compared with the analysis location of Ti-thermometry results published in ref. 17. Grains were further considered pristine if U/Pb concordance (that is, $^{206}\text{Pb}^*/^{238}\text{U age} / [^{207}\text{Pb}/^{206}\text{Pb}^* \text{ age}] \times 100$, expressed as percentage concordance) was, on average, 94%. The oxygen fugacities for these pristine grains are plotted in Fig. 2, and the data used are included in Supplementary Table 5. All lunar zircons with REE and Ti data from ref. 19 are included; grains reported in that study are ~94% concordant or better. The ~800,000-year-old zircons from the Bishop Tuff and those collected near present-day active mid-ocean ridges are all assumed to be pristine; all data with Ti-thermometry and REE data are included in Fig. 2b. Ce anomalies are calculated from La, Ce and Pr concentrations using equation (1) in the text but with chondrite normalized values; results agree with those reported in the above studies with the exception of ref. 30.

31. Trail, D. Watson, E. B. & Thomas, J. B. The incorporation of OH into zircon. *Am. Mineral.* **96**, 60–67 (2011).
32. Watson, E. B., Wark, D. A. & Thomas, J. B. Crystallization thermometers for zircon and rutile. *Contrib. Mineral. Petrol.* **151**, 413–433 (2006).
33. Ayers, J. C., Brenan, J. B., Watson, E. B., Wark, D. A. & Minarik, W. G. A new capsule technique for hydrothermal experiments using the piston cylinder apparatus. *Am. Mineral.* **77**, 1080–1086 (1992).

Subduction dynamics and the origin of Andean orogeny and the Bolivian orocline

F. A. Capitano¹, C. Faccenna², S. Zlotnik^{1,3,4} & D. R. Stegman⁵

The building of the Andes results from the subduction of the oceanic Nazca plate underneath the South American continent^{1,2}. However, how and why the Andes and their curvature, the Bolivian orocline, formed in the Cenozoic era (65.5 million years (Myr) ago to present), despite subduction continuing since the Mesozoic era³ (251.0–65.5 Myr ago), is still unknown. Three-dimensional numerical subduction models demonstrate that variations in slab thickness, arising from the Nazca plate's age at the trench, produce a cordilleran morphology consistent with that observed^{1,2}. The age-dependent sinking of the slab in the mantle drives traction towards the trench at the base of the upper plate, causing it to thicken. Thus, subducting older Nazca plate below the Central Andes can explain the locally thickened crust and higher elevations. Here we demonstrate that resultant thickening of the South American plate modifies both shear force gradients and migration rates along the trench to produce a concave margin that matches the Bolivian orocline. Additionally, the varying forcing along the margin allows stress belts to form in the upper-plate interior, explaining the widening of the Central Andes and the different tectonic styles found on their margins, the Eastern and Western Cordilleras². The rise of the Central Andes and orocline formation are directly related to the local increase of Nazca plate age and an age distribution along the margin similar to that found today; the onset of these conditions only occurred in the Eocene epoch⁴. This may explain the enigmatic delay of the Andean orogeny, that is, the formation of the modern Andes.

Several outstanding issues concerning the Andean orogeny currently remain unaddressed. Despite subduction and Western Cordillera tectonics beginning as early as the Cretaceous period³, ~140 Myr ago, widening to the Eastern Cordillera and Bolivian orocline formation did not start until the Eocene⁵, ~45 Myr ago. The plateau region sandwiched between the Cordilleras remained undeformed until the Miocene epoch⁶, ~10 Myr ago, when the whole Central Andes rose to reach modern elevations. Yet, it remains unexplained why the Andes and the orocline formed only in the Cenozoic, why the deformation then leaped 500 km into the plate's interior to the Eastern Cordillera, and why and how it affected the plateau region much later. Furthermore, an explanation is still lacking for the curvature of the Bolivian orocline, one of the few concave convergent margins on Earth.

The Andean chain displays distinct topographic features with a strong symmetry around the Central Andes^{7,8} (Fig. 1a). This segment of the belt is the widest, ~600–800 km, and hosts the highest elevations, ~5 km (Fig. 1b). Northwards and southwards, the altitude and width of the chain decrease smoothly. The Central Andes comprise the Western and the Eastern Cordilleras, both of similar elevation, separated by the Altiplano-Puna plateau (Fig. 1b, c). Crustal thicknesses of 60–80 km are measured in this region, whereas values of 40 km are found elsewhere⁹. The Central Andes corresponds to the largest curvature of the Bolivian orocline, marking a bend in the mountain chain.

The symmetry along the Andes correlates remarkably with the symmetric features of the Nazca plate^{7,8}. At the centre of this oceanic

plate, below the Central Andes, the lithosphere is oldest (~50 Myr), thickest, and most negatively buoyant, and progressively becomes younger towards the north and south, where mid-ocean ridges are subducting (Fig. 1a, b). Below the Central Andes, the dip of the Nazca slab shallows to ~50°. Two segments of very shallow slab border the Central Andes, but at depths greater than 200 km the slab bends again to dips of 70–80° (refs 4, 10). Faster subduction hinge migration occurs along the limbs of the continent¹¹. When extrapolated to a common time interval, geodetic and palaeomagnetic data on

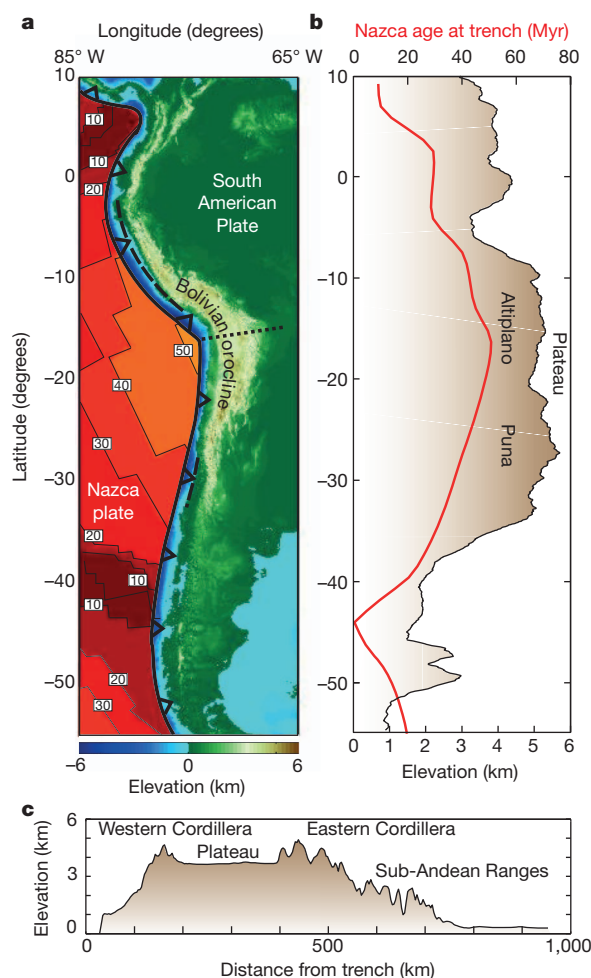


Figure 1 | The Andes and South American subduction. **a**, Andes topography (digital elevation model GTOPO30) and Nazca oceanic plate age⁴. The trench is indicated by thick line with triangles (dashed line for flat slab segments¹⁰). Dotted line, profile in **c**. **b**, Maximum along-margin Andes topography (black line with brown shading), and Nazca plate age at trench (red line) in Myr. **c**, Topographic profile across Central Andes. See **a** for location.

¹School of Geosciences, Monash University, Clayton, 3800 Victoria, Australia. ²Dipartimento di Scienze Geologiche, Università Roma Tre, 00146 Rome, Italy. ³School of Mathematical Sciences, Monash University, Clayton, 3800 Victoria, Australia. ⁴Laboratori de Calcul Numeric, Departament de Matemàtica Aplicada III, UPC-Barcelona Tech, Jordi Girona 1-3 E-08034 Barcelona, Spain. ⁵Scripps Institution of Oceanography, University of California San Diego, California 92093-0220, USA.

both flanks of the Bolivian orocline are similar, showing that the present-day margin motions and deformations are representative of the long-term growth and bending of the Andes^{5,12}.

The strong correlation between topography and extent of the Andean Cordillera with corresponding variations in the Nazca slab has long suggested a causal relationship with the underlying subduction dynamics^{7,8}. Several models have investigated how processes related to subduction can shape the Andes. Shear stress at the plates' interface¹³, possibly due to friction^{14–16}, sub-lithospheric mantle flow^{17,18}, or a combination of the two¹⁹, can reproduce the topography of the chain. Alternatively, relative plate convergence²⁰, lithospheric weakening¹⁵ and climate-driven surface processes^{14,21} have been invoked to explain the Andean evolution. However, the models proposed are unable to describe the evolution of the Andes and the Bolivian orocline as emerging features of a dynamic subduction system.

A regional-scale three-dimensional geodynamic model can help address these questions, by testing the relative contributions of the forces driving the motions and deformations around the convergent margin. Here we use a numerical model of subduction²², in which an overriding plate is coupled to a 6,000-km-wide plate subducting into the mantle under the pull of its own negative buoyancy, and plate motions and tectonic stresses are emergent quantities (Supplementary Information). We find that consideration of the third dimension is the key to understanding the role of strong lateral stress gradients

due to the heterogeneities in the subducting Nazca plate age and the South American plate thickness, along the convergent margin. In the downgoing lithosphere, age-dependent thickness influences the buoyancy, the largest driving force of subduction, and strength of the plate, controlling the dips, whereas in the upper plate, thickness affects the strength as well as the area of the margin interface, and thereby the integrated shear force locally resists subduction. Therefore, the subducting plate is modelled with either uniform thickness, or with a central thickened region, corresponding to the oldest oceanic lithosphere in the Nazca plate. Similarly, upper-plate model thickness is either uniform (40 km), or includes a 60–80 km thickened central portion, similar to the estimated crustal thicknesses along the South American margin⁹.

Results indicate, as expected, that the age of the downgoing lithosphere has a first-order control on the velocity of subducting plate and on the topography of the overriding plate. The larger negative buoyancy force of an older and thicker subducted slab drives faster convergence. Stress at convergent plate margins directly correlates to the slab vertical sinking rate, modulated by the dependence of buoyancy on age. This is because faster sinking drives more vigorous flow in the mantle wedge, increasing traction at the base of the overriding plate and causing it to thicken. In the models, this is indicated by vertical stresses and is a measure of the excess crust required to balance the stress, which is directly reflected in increased topography.

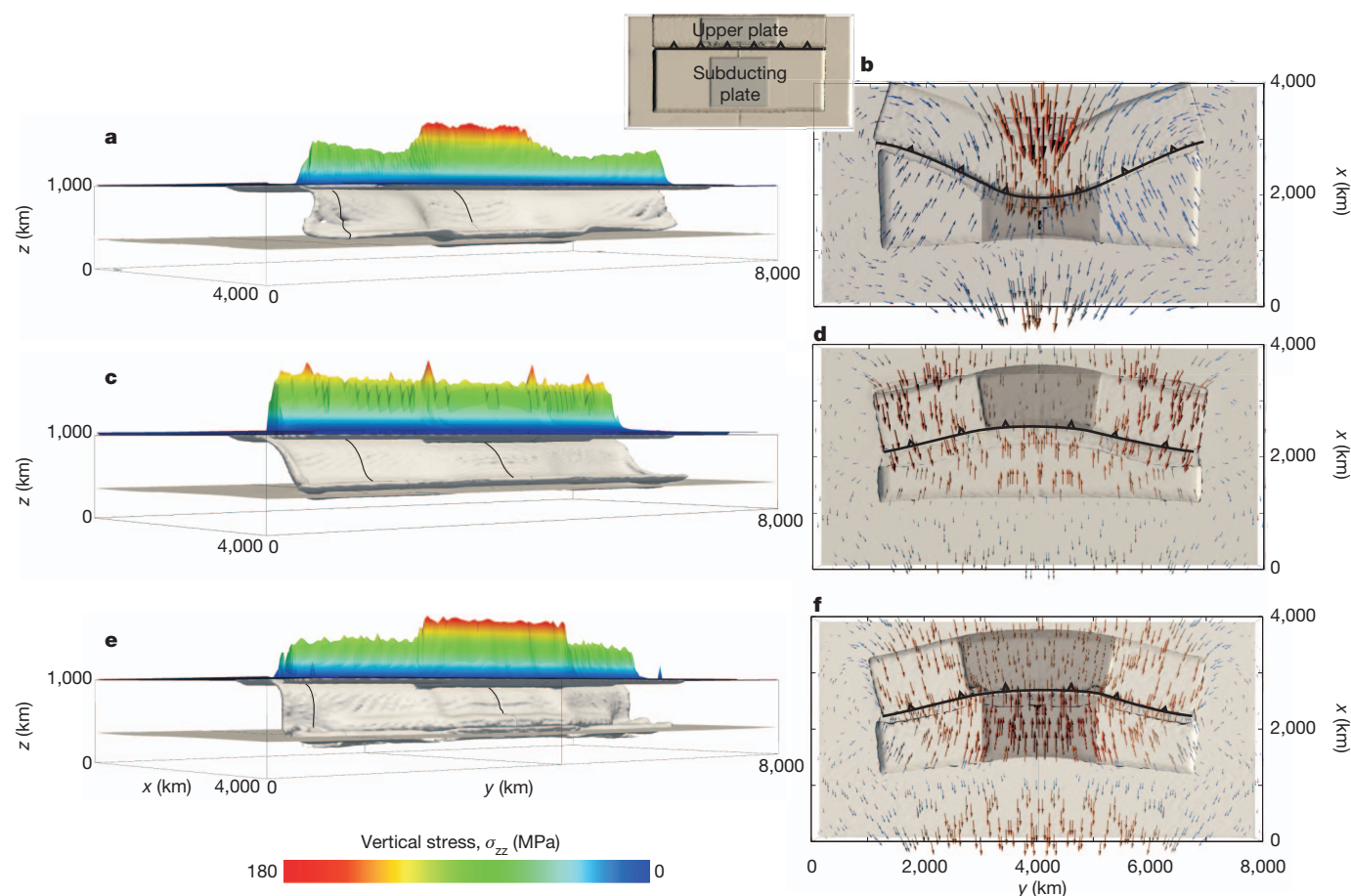


Figure 2 | Subduction models with slab morphologies, vertical stresses and velocities, after 2,000 km of subduction. Left column, three-dimensional view of the viscosity field contour (grey surface) and the vertical stress in the upper plate above the trench (coloured surface above the model box). The model space is a Cartesian box of $4,000 \times 8,000 \times 1,000$ km respectively in length (x), width (y) and height (z). Shown are the contour surface for the lower mantle's top (at $z = 340$ km), the subducting slab (between depths of 1,000 and 340 km) and plates at the surface (dark grey, $z = 1,000$ km). The black line at

$z = 1,000$ km is the top surface of the model. Right column, two-dimensional plan view of the surface, showing viscosity field contour (grey surface) and velocity field (arrows). **a, b**, Model with subducting-plate thickness heterogeneity. **c, d**, Model with upper-plate thickness heterogeneity. **e, f**, Model with thickness heterogeneities in both subducting and upper plates. Solid line with triangles, trench trace; dark shaded area, the thicker portion of the plates. Velocity is 0 to 8 cm yr^{-1} (blue to red). Top inset, initial configuration for **e, f**. Plates are initially straight and change morphology during subduction.

First, we tested the role that the older, thicker segment in the downgoing lithosphere has in subduction. This model includes a simplified version of the present-day Nazca plate age gradient at the trench, and generates several similarities with the Andean topography and Nazca subducted plate morphology along the margin. In the model's uniformly thick upper plate, vertical stresses reach a maximum of ~ 150 MPa (Fig. 2a), which are compensated by $\sim 5,500$ m of constant density continental crust (2.75 kg m^{-3}), and are thus compatible with Central Andean elevations. Also, vertical stresses decrease northwards and southwards, comparable to the observed elevation drop. Dip of the Nazca plate below the Central Andes, where oceanic lithosphere is older, is consistent with our model, as well as the steeper deep slab outwards, where the Nazca plate is younger. No flat slab occurs in our models. It is likely that other mechanisms become relevant at smaller scales²³, although these cannot hamper the deeper sinking of the slab. Faster subduction where the integrated slab pull is larger also promotes faster rollback, and the trench eventually forms a convex margin (Fig. 2b), opposite to the curvature of the Bolivian orocline.

The second model had a uniform subducting plate but included variation of the upper-plate thickness. In contrast to the first model, it did not reproduce north–south variation of vertical stress above the trench, but showed constant topography instead (Fig. 2c). This is related to constant velocities and slab dips (Fig. 2d), resulting from constant slab buoyancy and strength. In this model, upper-plate thickness variation along the trench reproduces realistic shear force gradients along the South American margin, increasing by a factor of ~ 2 below the Central Andes¹⁴ with respect to values further out. These gradients cause rollback to be slowest in the central, thickest part, ultimately causing bending in a concave morphology (Fig. 2d). This process might explain the long-term bending of the Bolivian orocline around the thicker Central Andean lithosphere.

When these heterogeneities are combined in a single model (Fig. 2e, f), the dynamics of the Nazca plate, the Andean topography and Bolivian orocline can all be reconciled. Results demonstrate that symmetric changes in the age of the subducting plate, older and thicker in its centre, cause symmetric changes in plate motions, dips and vertical stresses, reflected as a topographic high in the Central Andes, decreasing symmetrically outwards. In turn, the symmetric variations in thickness of the upper plate overcomes its natural tendency to bend into a convex margin²⁴, and instead forms a concave margin like the Bolivian orocline (Fig. 2e, f).

These different mechanisms, related to the subduction dynamics, can offer an explanation for the complexities in the Central Andean tectonics. Isostatic equilibrium of a uniform felsic 70–80 km crust⁹, mostly thickened with negligible thrusting²⁵, can account for the Western Cordillera elevations². Similar heights are explained by isostatic compensation of modelled vertical stresses above the trench. Instead, in the Eastern Cordillera, large crustal thickness is achieved by shortening along deep-seated thrusts²⁵, thus requiring stress propagation far into the continent's interior. Figure 3 shows intraplate stress and dynamic pressure in the upper plate for the model that includes both heterogeneities. Two zones of high shear stress develop at an angle to the trench (Fig. 3a), above the thickest segment of the downgoing plate. Along these belts the large horizontal stresses accommodate the different forcing at the margin. The distinct spatial localization of these mechanisms leaves an area in between where stress is negligible (Fig. 3c). This may explain the different regimes underlying the Western and the Eastern Cordilleras and their morphology, as well as the lack of deformation in the plateau during the early stages of Central Andes evolution. Additionally, the pressure gradient towards the centre of the upper plate might drive orogeny-parallel crustal flow towards the plateau²⁶.

Further deformation only occurred subsequently (post-Oligocene), migrating to the plateau and the outer sub-Andean ranges²⁷. The results from three models where the thickness of the upper plate was uniform at 40, 60 or 80 km, show that thicker upper plates are stronger and

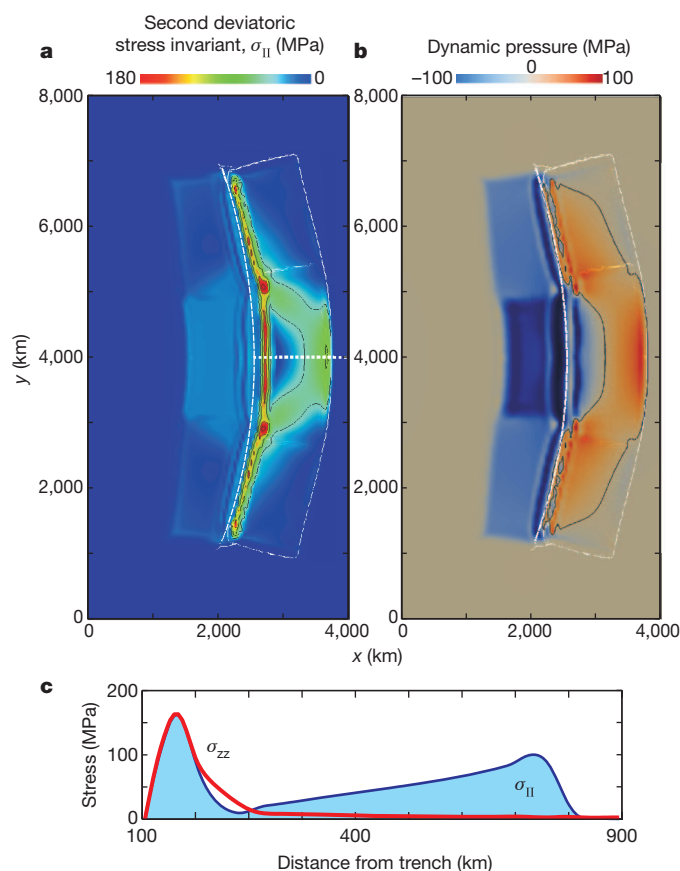


Figure 3 | Surface stress and pressure from the model with heterogeneities in both plates. This model shown in Fig. 2e, f; the upper plate is outlined in white. **a**, Second deviatoric stress invariant, σ_{II} . High stress above the trench and across the upper plate isolates a low-stress area. **b**, Dynamic pressure. Negative pressure occurs above the denser subducting plate segment. Black line, contour of zero pressure. **c**, Upper-plate stress, cross-section. Large vertical stress (σ_{zz} , red line) above the trench indicates thickening; high stress invariant (σ_{II} , blue line with shading) in the plate interiors and zero vertical stress indicates horizontal major stress. Profile location, dashed line in **a**.

compression propagates further into the plate's interior (Fig. 4a), which is required for outward orogen growth. Compression in the plateau is accommodated by thrust tectonics in the Bolivian Altiplano⁶, as opposed to pure-shear thickening in the Argentinean Puna⁶, achieved by diffuse faulting. In the Puna, the thickened lithosphere has eventually delaminated²⁸. In the models, we found that the peak differential stress deepens with the thickness of the modelled upper plate (Fig. 4b). Thus, we argue that similar stress localization at subcrustal depths might have triggered the delamination of the South American lithosphere. Stress focusing occurs when the upper-plate model thickens uniformly, but this might not be the case if plastic deformation continuously accommodates compression. This possibly explains why the lithosphere has only delaminated under some parts of the plateau⁹, even if it has similarly thickened and become unstable throughout.

Our models suggest that such diverse evolution can be reconciled with the boundary conditions that the Nazca slab buoyancy imposed on the South American continent. So the age of the subducted lithosphere going back in time might provide a first-order estimate of the driving forces during the Cenozoic, which could be verified against the timing of the Andean orogeny. The growth of both Cordilleras above the trench requires an increase in the age of the subducting plate, which would also drive a surge in the convergence. In contrast, the widening of the central part of the orogen and the bending of the orocline are both a response to the age gradients along the trench, leading to heterogeneous thickening of the upper plate. The increase of subducting plate age and the strong along-trench age gradients are

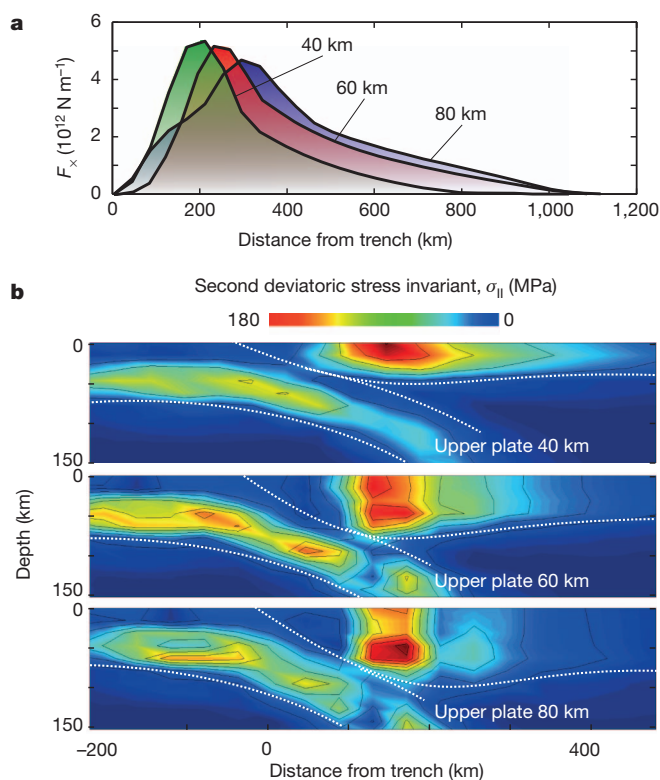


Figure 4 | Effects of upper-plate thickening. **a**, Integrated horizontal compressive force, F_x , propagated from trench in the upper plate. Models shown have overriding plates 40, 60 and 80 km thick. Compression propagates farther from trench with thicker, stiffer plates. **b**, Cross-section of the second deviatoric stress invariant. Profile location, dashed line in Fig. 3a.

conditions that might not occur at the same time, nor might they correlate, thus it is a natural (not exceptional) consequence that orogen growth occurred long after subduction had started.

Reconstructions of the Nazca plate age⁴ for the past 60 Myr (Supplementary Data) show that a strong, along-trench age gradient formed ~ 45 – 40 Myr ago, between 30° and 45° S (ref. 4), where older subducting plate entered the trench. The increased driving force can explain the fast convergence during this period^{4,20} and the inception of Eastern Cordillera tectonics²⁷. The present-day age gradient was not established until ~ 20 Myr ago⁴, when the Nazca plate migrated to its present-day location. This positioned ~ 70 -Myr-old lithosphere offshore from South America, which then became the oldest lithosphere subducted under the Central Andes since the early Cenozoic. This accounts for the second surge in the convergence rate^{4,20} as well as the local growth of the topography, above the northern and southern average elevation²⁷. As prescribed by our models, the timing of crustal growth matches the two major phases of oroclinal bending⁵. The progressive regional reduction in the age of the subducting Nazca plate since 15 Myr ago explains the decline of convergence rates to present-day values. The current convergence velocity varies little along the South American trench, reaching a maximum further south than the Central Andes²⁹, at 35 – 40° S. This is possibly due to the increased pull of the deeper slab at these latitudes, where older, denser lithosphere has subducted before 25 Myr ago. Therefore, variations in the dominant plate driving force, slab pull, that naturally arise from heterogeneities in the seafloor age of oceanic plates, are linked to intervals of intense tectonic activity and quiescence, and to the corresponding time-history of convergent motions.

Heterogeneities in the downgoing and upper plates are features commonly found along all convergent margins. Thus, their controls on Andean-type chains, plate margin morphologies and continental deformation must have a global relevance.

Received 10 November 2010; accepted 26 September 2011.

Published online 23 November 2011.

- Isacks, B. L. Uplift of the central Andean plateau and bending of the Bolivian orocline. *J. Geophys. Res.* **93**, 3211–3231 (1988).
- Allmendinger, R. W., Jordan, T. E., Kay, S. M. & Isacks, B. L. The evolution of the Altiplano-Puna plateau of the central Andes. *Annu. Rev. Earth Planet. Sci.* **25**, 139–174 (1997).
- Mpodozis, C. & Ramos, V. A. in *Geology of the Andes and its Relation to Hydrocarbon and Mineral Resources* Vol. 11 (eds Erickson, G. E., Cañas Pinochet, M. T. & Reinemund, J. A.) 59–90 (1990).
- Sdrolias, M. & Müller, R. D. Controls on back-arc basin formation. *Geochem. Geophys. Geosyst.* **7**, Q04016, doi:10.1029/2005GC001090 (2006).
- Arriagada, C., Roperch, P., Mpodozis, C. & Cobbold, P. Paleogene building of the Bolivian Orocline: tectonic restoration of the central Andes in 2-D map view. *Tectonics* **27**, TC6014, doi:10.1029/2008TC002269 (2008).
- Allmendinger, R. W. & Gubbels, T. Pure and simple shear plateau uplift, Altiplano-Puna, Argentina and Bolivia. *Tectonophysics* **259**, 1–13 (1996).
- Gephart, J. W. Topography and subduction geometry in the central Andes: clues to the mechanics of a noncollisional orogen. *J. Geophys. Res.* **99** (B6), 12279–12288 (1994).
- Jordan, T. E. *et al.* Andean tectonics related to geometry of subducted Nazca plate. *Geol. Soc. Am. Bull.* **94**, 341–361 (1983).
- Beck, S. L. & Zandt, G. The nature of orogenic crust in the central Andes. *J. Geophys. Res.* **107** (B10), 2230, doi:10.1029/2000JB000124 (2002).
- Gutscher, M. A., Spakman, W., Bijwaard, H. & Engdahl, E. R. Geodynamics of flat subduction: seismicity and tomographic constraints from the Andean margin. *Tectonics* **19**, 814–833 (2000).
- Schellart, W. P. *et al.* Evolution and diversity of subduction zones controlled by slab width. *Nature* **446**, 308–311 (2007).
- Allmendinger, R. W. *et al.* Bending the Bolivian orocline in real time. *Geology* **33**, 905–908 (2005).
- Yáñez, G. & Cembrano, J. Role of viscous plate coupling in the late Tertiary Andean tectonics. *J. Geophys. Res.* **109**, B02407, doi:10.1029/2003JB002494 (2004).
- Lamb, S. & Davis, P. Cenozoic climate change as a possible cause for the rise of the Andes. *Nature* **425**, 792–797 (2003).
- Sobolev, S. V. & Babeyko, A. Y. What drives orogeny in the Andes? *Geology* **33**, 617–620 (2005).
- Iaffaldano, G. & Bunge, H. P. Strong plate coupling along the Nazca–South America convergent margin. *Geology* **36**, 443–446 (2008).
- Russo, R. M. & Silver, P. G. Cordillera formation, mantle dynamics, and the Wilson cycle. *Geology* **24**, 511–514 (1996).
- Wdowinski, S., O'Connell, R. J. & England, P. A continuum model of continental deformation above subduction zones: application to the Andes and the Aegean. *J. Geophys. Res.* **94** (B8), 10331–10346 (1989).
- Husson, L. & Ricard, Y. Stress balance above subduction: application to the Andes. *Earth Planet. Sci. Lett.* **222**, 1037–1050 (2004).
- Pardo-Casas, F. & Molnar, P. Relative motions of the Nazca (Farallon) and South American plates since late Cretaceous time. *Tectonics* **6**, 233–248 (1987).
- Montgomery, D. R., Balco, G. & Willett, S. D. Climate, tectonics, and the morphology of the Andes. *Geology* **29**, 579–582 (2001).
- Capitanio, F. A., Stegman, D. R., Moresi, L. & Sharples, W. Upper plate controls on deep subduction, trench migrations and deformations at convergent margins. *Tectonophysics* **483**, 80–92 (2010).
- van Hunen, J., van der Berg, A. P. & Vlaar, N. J. On the role of subducting oceanic plateaus in the development of shallow flat subduction. *Tectonophysics* **352**, 317–333 (2002).
- Morra, G., Regenauer-Lieb, K. & Giardini, D. On the curvature of oceanic arcs. *Geology* **34**, 877–880 (2006).
- Kley, J. & Monaldi, C. R. Tectonic shortening and crustal thickness in the Central Andes: how good is the correlation? *Geology* **26**, 723–726 (1998).
- Gerbault, M., Martinod, J. & Hérail, G. Possible orogeny-parallel lower crustal flow and thickening in the Central Andes. *Tectonophysics* **399**, 59–72 (2005).
- Elger, K., Oncken, O. & Glodny, J. Plateau-style accumulation of deformation: Southern Altiplano. *Tectonics* **24**, TC4020, doi:10.1029/2004TC001675 (2005).
- Garzone, C. N., Molnar, P., Libarkin, J. C. & MacFadden, B. J. Rapid late Miocene rise of the Bolivian Altiplano: evidence for removal of mantle lithosphere. *Earth Planet. Sci. Lett.* **241**, 543–556 (2006).
- DeMets, C., Gordon, R. G. & Argus, D. F. Geologically current plate motions. *Geophys. J. Int.* **181**, 1–80 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Faccenda, R. F. Weinberg and J. P. Brun for discussions. F.A.C. was supported by the Australian Research Council's Discovery Projects DP0987374. D.R.S. was supported in part by the G. Unger Vetlesen Foundation.

Author Contributions F.A.C. designed and performed the numerical modelling. F.A.C. and S.Z. processed the models' output and the data. All authors contributed to the writing of the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to F.A.C. (fabio.capitanio@monash.edu).

Correlation between deep fluids, tremor and creep along the central San Andreas fault

Michael Becken^{1,2,†}, Oliver Ritter¹, Paul A. Bedrosian³ & Ute Weckmann^{1,2}

The seismicity pattern along the San Andreas fault near Parkfield and Cholame, California, varies distinctly over a length of only fifty kilometres. Within the brittle crust, the presence of frictionally weak minerals, fault-weakening high fluid pressures and chemical weakening are considered possible causes of an anomalously weak fault northwest of Parkfield^{1–4}. Non-volcanic tremor from lower-crustal and upper-mantle depths^{5–7} is most pronounced about thirty kilometres southeast of Parkfield and is thought to be associated with high pore-fluid pressures at depth⁸. Here we present geophysical evidence of fluids migrating into the creeping section of the San Andreas fault that seem to originate in the region of the uppermost mantle that also stimulates tremor, and evidence that along-strike variations in tremor activity and amplitude are related to strength variations in the lower crust and upper mantle. Interconnected fluids can explain a deep zone of anomalously low electrical resistivity that has been imaged by magnetotelluric data southwest of the Parkfield–Cholame segment. Near Cholame, where fluids seem to be trapped below a high-resistivity cap, tremor concentrates adjacent to the inferred fluids within a mechanically strong zone of high resistivity. By contrast, sub-vertical zones of low resistivity breach the entire crust near the drill hole of the San Andreas Fault Observatory at Depth, northwest of Parkfield, and imply pathways for deep fluids into the eastern fault block, coincident with a mechanically weak crust and the lower tremor amplitudes in the lower crust. Fluid influx to the fault system is consistent with hypotheses of fault-weakening high fluid pressures in the brittle crust.

Magnetotelluric data sense the electrical resistivity of Earth, a physical parameter that is particularly sensitive to the presence of low-resistivity phases such as aqueous fluids, partial melts or metallic compounds. Fluid phases have electrical resistivities orders of magnitude lower than that of the rock matrix, and relatively small amount of fluids, when interconnected, can thus decrease bulk rock resistivity by several orders of magnitude⁹. Fluids additionally have a notable weakening effect on the rheology of rocks, even more so if fluids form an interconnected network¹⁰. Measurements of electrical resistivity can therefore be used to constrain the volume of subsurface fluids, their interconnectivity and the rheology of the crust and mantle.

We collected magnetotelluric data along seven profiles across the San Andreas fault (SAF) near Parkfield and Cholame, covering the tremor concentration zone near Cholame and the region of transition from locked to creeping behaviour (Fig. 1a; data maps are provided in Supplementary Fig. 1). The most prominent structure revealed by the magnetotelluric data is a deep low-resistivity zone (1–5 Ω m) centred 30–40 km southwest of the SAF at a depth of more than 15–20 km (Fig. 1b). This anomaly occupies a broad region in the lower crust and upper mantle between the surface traces of the Rinconada fault (a former strand of the SAF) and the modern-day SAF. Along profiles 4–7 (Fig. 1b), which cross the SAF in the tremor concentration zone near the northern end of the locked segment^{5,7}, crustal resistivities are

in excess of 500 Ω m for the Pacific plate, by contrast with the moderately resistive crustal formations (10–100 Ω m) of the North American plate. These values are typical for the granitic rocks of the Salinian block and for the metamorphic and sedimentary sequences contained within the Franciscan formation, respectively. The subduction complex constituting the Franciscan formation includes a great variety of lithologies, including fragments of oceanic crust, bodies of greenstones and other metamorphic and volcanic rocks. Along-strike variations in the resistivity structure of the Franciscan formation, such as the high resistivities northeast of the SAF near Parkfield (Fig. 1b, profiles 2 and 3), reflect this heterogeneity.

The SAF itself appears most prominently beneath the Cholame valley (Fig. 1b, profile 5, and Fig. 2b) as a vertical resistivity contrast extending below the brittle–ductile transition to depths of 20 km or more. To the northwest, where the SAF shows creep and microseismicity, profiles 1–3 image a low-resistivity (5–30 Ω m) linkage between the deep, low-resistivity zone to the southwest and the SAF to the northeast (Fig. 2a), by marked contrast with the resistivity models farther south.

The anomalously low electrical resistivity in the lower crust and upper mantle provides geophysical evidence for a deeply seated fluid source offset to the southwest of the SAF. Low electrical resistivity has been reported at similar depths along numerous active margins, including the Pacific–Australian plate collision zone in New Zealand¹¹, that beneath the Tibetan plateau¹² and the altiplano of the Andes in Chile and Bolivia¹³. Explanations involved the presence of partial melt¹⁴ in combination with aqueous fluids, imply a weakened rheology and support models of lower-crustal flow¹². Modelling of the California Coast Ranges heat flow anomaly suggests that temperatures reach ~ 800 °C at a depth of 30 km near Parkfield¹⁵; this is too cold for the upper mantle to be partially molten. The resistivity of aqueous fluids depends strongly on temperature and salinity. The salinity of sea water is a conservative assumption for the fluid phase. At 800 °C, the electrical resistivity of sea water is 0.03 Ω m, which is approximately ten times lower than the value at room temperature¹⁶. The Hashin–Shtrikman upper bound¹⁷ for a two-phase system composed of a moderately resistive rock matrix (100 Ω m) and a low-resistivity aqueous fluid phase (0.03 Ω m) requires 3 vol% of perfectly interconnected porosity over a large volume to explain bulk resistivities of 1 Ω m; higher fluid salinities would require less interconnection and/or lower porosity.

Aqueous fluid networks can explain the observed low resistivities, but not unambiguously. A region in the crust weakened by shearing and faulting can become mineralized by chemical precipitation reactions^{18,19}. Continued shearing deformation can further increase the interconnectivity of such mineralized zones and result in low electrical resistivities²⁰. However, independent geophysical observations strongly support our preferred interpretation of a deep fluid source. A seismic refraction profile approximately coincident with profile 5 (Fig. 1b) reveals a low-velocity zone with compressional wave speeds of less than 5 km s^{–1} at a depth of 15–22 km (ref. 21), coincident with the top of the low-resistivity zone and consistent with the presence of fluids in this region (Fig. 2b).

¹GFZ German Research Centre for Geosciences, Telegrafenberg, 14473 Potsdam, Germany. ²University Potsdam, Institute of Geosciences, Karl-Liebknecht-Strasse 24, 14476 Potsdam-Golm, Germany.

³US Geological Survey, MS 964, Box 25046, Building 20, Denver, Colorado 80225, USA. [†]Present address: Westfälische Wilhelms Universität Münster, Institute of Geophysics, Corrensstrasse 24, 48149 Münster, Germany.

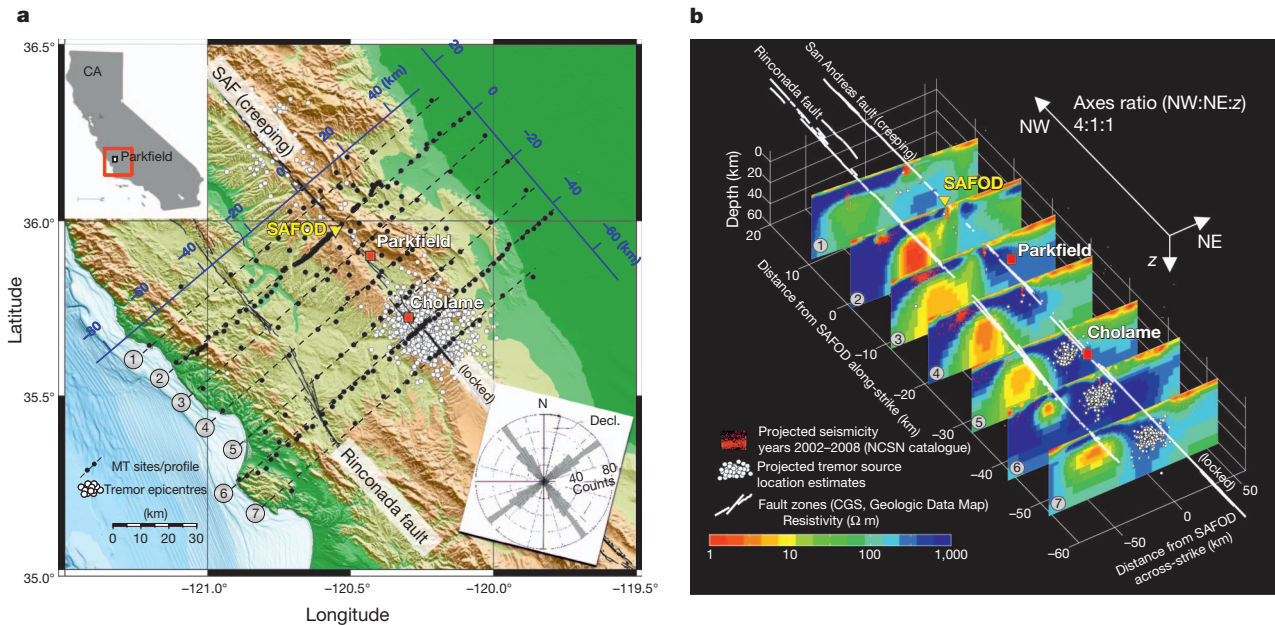


Figure 1 | Magnetotelluric resistivity sections across the central San Andreas fault. **a**, Map view. Black dots show magnetotelluric stations along seven ~130-km-long profiles. Magnetotelluric (MT) tensor decomposition³⁰ yields geoelectric strike estimates (inset rose diagram; decl., declination) that are consistent with the orientation of the SAF. SAF creep rates decay from ~30 mm yr⁻¹ northwest of SAFOD to less than 5 mm yr⁻¹ southeast of Cholame⁴. White dots show tremor epicentres⁷. The blue axes show coordinates relative to SAFOD (used in **b** and in Figs 2 and 3). **b**, Electrical

Fluids in this depth range are also likely to be involved in the generation of deep, non-volcanic tremor along this fault segment^{6,8}. Teleseismic and tidal triggering of tremor episodes at the SAF indicate that tremor nucleates in regions close to mechanical failure at near-lithostatic fluid pressures⁸. For the magnetotelluric profiles 5–7 across the tremor concentration zone near Cholame, the inversion reveals a deep, low-resistivity region that is bounded from above and laterally by resistive formations (~1,000 Ω m; Figs 1b and 2b). Here the tremor source region seems to coincide with the high-resistivity rocks, adjacent to the less resistive region that is a potential source of fluid. Inversions of the magnetotelluric data confirm that the high-resistivity cap is a very robust feature of the models. Any low-resistivity connection to the upper crust (as observed northeast of Parkfield) results in a significant increase in data mis-fit (Supplementary Information and Supplementary Fig. 6).

Tremor near Cholame separates spatially into a southwestern zone with periodic tremor episodes and a northeastern zone with aperiodic episodes⁷. Our resistivity models suggest that the fluid source (that is, the low-resistivity region) is located southwest of the periodic tremor zone. Ongoing fluid generation by mantle dehydration reactions⁴ could result in high fluid pressures in the low-resistivity zone and, consequently, in fluids being continuously driven through fracture systems towards and into the more resistive tremor regions. Observations of high seismic reflectivity²¹ (Fig. 2b) at tremor source depths between the low-resistivity fluid source and the more resistive tremor zone are consistent with deformed and fractured material. Lateral migration of fluids could be responsible for increased fluid pressures in the tremor source region⁸. The more periodic tremor episode occurrences are located closer to the fluid source and could reflect a failure mechanism involving cyclic fluid pressure accumulation and release.

A similar fluid-triggering model has been proposed to account for the globally observed correlation between electrical resistivity and brittle failure in the upper crust²², including along the SAF²³. Brittle rock failure and earthquakes often occur close to the boundary of, or within, regions of high electrical resistivity, but adjacent to zones of low

resistivity sections. Low resistivity in the lower crust and upper mantle is attributed to hot saline fluids. Fluid can upwell into the creeping SAF near SAFOD. Near the locked segment, fluids seem to be trapped adjacent to the tremor concentration zone and to be prevented from escaping into the upper-crustal portion of the SAF. Tremor sources⁷ (white dots) and seismicity (red dots) are superimposed. NCSN, Northern California Seismic Network; CGS, California Geological Survey.

resistivity²². These observations have been explained in terms of the migration of fluids from a permeable, mechanically weak source into a less permeable, mechanically strong region of high electrical resistivity. This transport can induce high fluid pressures and rock failure within the high-resistivity region, depending on the concentration of stress and the penetration depth of fluids into the rocks. Hydrofracturing or reaction-enhanced permeability²⁴ can be efficient mechanisms to facilitate fluid flow within low-permeability rocks under persistent, high confining pressures in the lower crust and upper mantle.

Tremor amplitudes along the central SAF vary by a factor of seven⁶ and correlate with variations in resistivity by a factor of 100 between Cholame and northwest of Parkfield (Fig. 3). High-amplitude tremors seem to originate predominantly in the high-resistivity (~1,000 Ω m) lower-crustal formations beneath Cholame, whereas the less resistive regions (5–30 Ω m) northwest of Parkfield have low tremor amplitudes or are devoid of measurable tremor (Fig. 3a). Deformation at lower-crustal depth is generally ductile. However, high fluid pressures can facilitate brittle failure also within an otherwise ductile deformation regime^{6,8}. Under these circumstances, brittle failure rather than ductile deformation has been considered an appropriate model for tremor⁸, and tremor amplitude may be a key to illuminating the physical conditions within the source region. The apparent systematic spatial variations of tremor activity and amplitudes and electrical resistivity along the deep central SAF⁶ (Fig. 3b) probably reflect variable conditions for brittle failure and indicate changes in mechanical properties and hydraulic conditions along the deep extension of the fault zone. The correspondence between resistivity and tremor amplitude leads us to conjecture that strength controls tremor amplitudes. Figure 3c depicts electrical resistance, here defined as the inverse of the depth-integrated reciprocal resistivity, as a qualitative measure of the mechanical strength of the tremor zone. The resistivity models and the variations of lower-crustal resistance argue for mechanically stronger rocks in the high-resistivity tremor concentration zone beneath Cholame as opposed to mechanically weaker material in the less resistive lower crust northeast of Parkfield. Weak lower-crustal material near the San

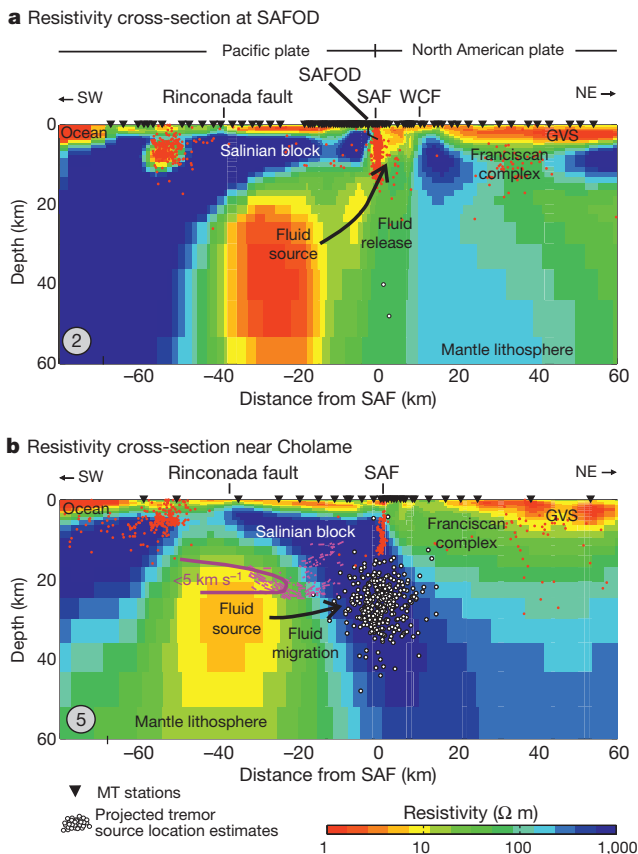


Figure 2 | Two distinct electrical resistivity models. **a**, Low-electrical-resistivity pathways for deep fluids breach the entire crust near SAFOD and provide a linkage (black arrow) between the fluid source beneath the Salinian block on the Pacific side and the upper crust on the North American side of the fault. Fluid flow across the plate boundary probably occurs at the base of the crust below the seismically defined fault (red dots, seismicity). **b**, Fluid pathways into the crust are absent from the Cholame segment of the fault. Here the low-resistivity fluid source is surrounded by highly resistive material and isolated from the SAF. Tremor (white dots) concentrates in the high-resistivity zone⁷. A wedge-shaped low-velocity zone (thick purple, 5 km s^{-1} seismic low-velocity contour interpreted from seismic refraction data²¹) coincides with the top of the low-resistivity fluid region. High seismic reflectivity (thin purple lines)²¹ between the low-resistivity/low-velocity zone and the tremor region indicates heterogeneous material, such as fracture or deformation zones, that in turn can provide pathways for lateral fluid flow. WCF, Waltham canyon fault; GVS, Great Valley sequence.

Andreas Fault Observatory at Depth (SAFOD) is also consistent with there being at a depth of 20–25 km an anisotropic, low-velocity layer with a high compressional wave/shear wave ratio²⁵. This is inferred from receiver functions, and the layer is interpreted as a sliver of serpentinite²⁵.

The weakness of the low-resistivity lower crust near SAFOD can be explained by a fluid-filled porosity network. Our resistivity models show that the deep fluid source is offset to the southwest of the SAF and located beneath the Pacific plate. From there, low-resistivity fluid pathways breach the lower crust near SAFOD and open out into an upper-crustal wedge of low electrical resistivity between the seismically defined SAF to the southwest and the Waltham canyon fault to the northeast²⁶. This crustal zone of low resistivity is less pronounced at profile 1 (Fig. 1b), and it is unknown whether it extends along the creeping segment.

Fluid ascent across the Pacific/North American plate boundary into the fault-adjacent country rock of the North American plate can contribute to super-hydrostatic fluid pressures along the creeping section of the SAF. Support for the existence of a connection to the lower crust and upper mantle comes from increased concentrations of

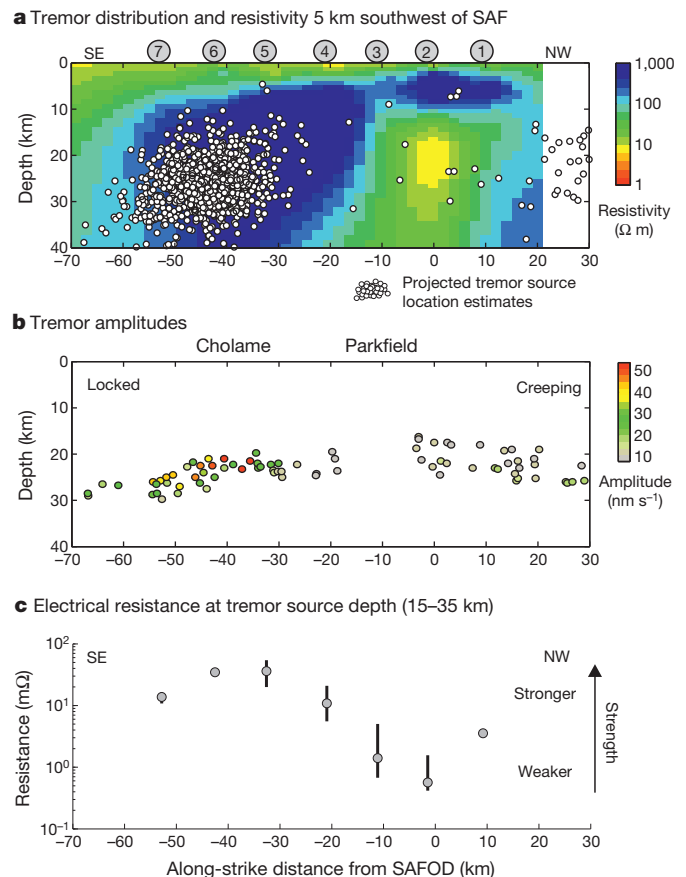


Figure 3 | Relation between electrical resistivity and tremor. **a**, Tremor⁷ (white dots) projected onto a resistivity slice, which is interpolated from the two-dimensional inversion models at a distance of -5 km from the SAF. **b**, Tremor amplitude estimates of individual tremor episodes⁶. **c**, Electrical resistance, the inverse of the depth-integrated reciprocal resistivity, computed for the tremor source depth range -5 km from the SAF; vertical bars indicate minimum and maximum resistances in the distance range $[-10 \text{ km}, 0 \text{ km}]$ along the individual profiles. Tremor concentrations and the largest tremor amplitudes correlate with the regions of high resistivity. By contrast, zones with low tremor activity and low tremor amplitudes are associated with areas of low resistivity. We interpret electrical resistance in terms of mechanical strength to explain tremor amplitude variations along the SAF.

mantle-derived gas found at SAFOD²⁷ and in nearby water wells penetrating the eastern fault block. Mantle-derived fluids contributing to fault-weakening high fluid pressures⁴ are considered one explanation for the frictional weakness of the SAF northeast of Parkfield. However, the role of fluids in the mechanics of this fault segment is still debated, because SAFOD observations did not provide evidence for increased pore pressures directly within the actively deforming fault strands at the depth where the SAFOD borehole intersects the fault²⁸.

Magnetotelluric data cannot provide definitive information about the nature of the fluid source in the lower crust and upper mantle. It has been suggested that fluids are released from ongoing dehydration of serpentinitized mantle wedge material⁴, a relic of former subduction along coastal California. Such dehydration can last for tens of millions of years⁴ and can be responsible for high fluid pressures in the tremor source region and, ultimately, the supply of fluids into the crust near SAFOD. We speculate that crustal pathways for fluids near SAFOD reflect zones of intensively deformed material that coincide with fault-related exhumation of serpentinite to shallower crustal levels. Fluid release from lower-crustal and upper-mantle formations can cause higher fluid pressures at shallower levels and result in mechanical weakening of the seismogenic SAF and the fault-adjacent country rocks of the North American Plate; support fluid-assisted chemical

weakening; and increase chemical reaction rates. To the southeast, where fluids are trapped beneath an impermeable cap in the crust, the fault is in a transitional-to-locked state. High pore pressures within the mantle and lower crust, accompanied by episodic fluid redistribution and seismic energy release seem to be responsible for the generation of tremor. Along the central SAF, the source of fluids migrating into the creeping section and stimulating the tremor region is offset to the southwest of the surface trace of the SAF. Tremor amplitude variations along the deep extension of the SAF seem to reflect fluid-controlled strength variations.

METHODS SUMMARY

The magnetotelluric method uses natural electromagnetic field variations to probe the electrical resistivity structure of Earth's interior. For this study, magnetotelluric data were collected in a series of field experiments between 2005 and 2008 using up to 45 simultaneously recording instruments. Robust remote and multi-remote reference processing was used to estimate magnetotelluric transfer functions. The data were inverted into electrical resistivity, using a regularized two-dimensional nonlinear conjugate gradient approach²⁹ (Supplementary Figs 2–8). Regularization of the model is necessary because the magnetotelluric method uses the diffusion of low-frequency natural electromagnetic fields within the conducting Earth, which results in a mathematically ill-posed, under-determined inverse problem. The inversion minimizes a Tikhonov-regularized functional as a trade-off between a structure penalty functional and the data residual norm weighted with the data variances. The structure penalty is a weighted integral of the squared Laplacian of the logarithmic model. Weights have been defined to generate increasing horizontal and vertical smoothing with increasing depth. Two-dimensional inversion assumes that resistivity is uniform along strike. This assumption is not strictly satisfied given the partly three-dimensional characteristics of the magnetotelluric data (Supplementary Information and Supplementary Fig. 9). However, the predominant magnetotelluric impedance strike directions determined from tensor decomposition³⁰ are consistent with a predominant geological strike that parallels the SAF in the region. A down-weighting scheme was applied to data with three-dimensional effects in to reduce bias in the two-dimensional inversion²⁶.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 7 November 2010; accepted 28 September 2011.

- Moore, D. E. & Rymer, M. J. Talc-bearing serpentinite and the creeping section of the San Andreas fault. *Nature* **448**, 795–797 (2007).
- Holdsworth, R. E. *et al.* Fault rocks from the SAFOD core samples: implications for weakening at shallow depths along the San Andreas Fault, California. *J. Struct. Geol.* **33**, 132–144 (2011).
- Lockner, D. A., Morrow, C., Moore, D. & Hickman, S. Low strength of deep San Andreas fault gouge from SAFOD core. *Nature* **472**, 82–85 (2011).
- Fulton, P. M. & Saffer, D. M. Potential role of mantle-derived fluids in weakening the San Andreas Fault. *J. Geophys. Res.* **114**, B07408 (2009).
- Nadeau, R. M. & Dolenc, D. Nonvolcanic tremors deep beneath the San Andreas Fault. *Science* **307**, 389 (2005).
- Shelly, D. R. & Hardebeck, J. L. Precise tremor source locations and amplitude variations along the lower-crustal central San Andreas Fault. *Geophys. Res. Lett.* **37**, L14301 (2010).
- Zhang, H., Nadeau, R. M. & Toksoz, M. N. Locating non-volcanic tremors beneath the San Andreas Fault using a station-pair double-difference location method. *Geophys. Res. Lett.* **37**, L13304 (2010).
- Thomas, A. M., Nadeau, R. M. & Bürgmann, R. Tremor-tide correlations and near-lithostatic pore pressure on the deep San Andreas fault. *Nature* **462**, 1048–1051 (2009).
- Guéguen, Y. & Palciauskas, V. *Introduction to the Physics of Rocks* 182–211 (Princeton Univ. Press, 1994).
- Bürgmann, R. & Dresen, G. Rheology of the lower crust and upper mantle: evidence from rock mechanics, geodesy, and field observations. *Annu. Rev. Earth Planet. Sci.* **36**, 531–567 (2008).
- Wannamaker, P. E. *et al.* Fluid and deformation regime of an advancing subduction zone at Marlborough, New Zealand. *Nature* **460**, 733–736 (2009).
- Unsworth, M. *et al.* Crustal rheology of the Himalaya and Southern Tibet inferred from magnetotelluric data. *Nature* **438**, 78–81 (2005).
- Brasse, H. *et al.* The Bolivian Altiplano conductivity anomaly. *J. Geophys. Res.* **107**, 2096 (2002).
- Schilling, F. R. & Partzsch, G. M. Quantifying partial melt fraction in the crust beneath the central Andes and the Tibetan plateau. *Phys. Chem. Earth* **26**, 239–246 (2001).
- Sass, J. H. *et al.* Thermal regime of the San Andreas Fault near Parkfield, California. *J. Geophys. Res.* **102**, 27575–27585 (1997).
- Quist, A. S. & Marschall, W. L. Electrical conductances of aqueous sodium chloride solutions from 0 to 800° and at pressures to 4000 bars. *J. Phys. Chem.* **72**, 684–703 (1968).
- Hashin, Z. & Shtrikman, S. A variational approach to the theory of effective magnetic permeability of multiphase materials. *J. Appl. Phys.* **33**, 3125–3131 (1962).
- Weckmann, U., Ritter, O. & Haak, V. A magnetotelluric study of the Damara Belt in Namibia: 2. MT phases over 90° reveal the internal structure of the Waterberg Fault/Omaruru Lineament. *Phys. Earth Planet. Inter.* **138**, 91–112 (2003).
- Ritter, O. *et al.* Electrical conductivity images of active and fossil fault zones. *Spec. Publ. Geol. Soc. (Lond.)* **245**, 165–186 (2005).
- Ritter, O. *et al.* Very high electrical conductivity beneath the Münchberg Gneiss area in Southern Germany: implications for horizontal transport along shear planes. *Geophys. J. Int.* **139**, 161–170 (1999).
- Trehu, A. M. & Wheeler, W. H. Possible evidence for subducted sedimentary materials beneath central California. *Geology* **15**, 254–258 (1987).
- Gürer, A. & Bayrak, M. Relation between electrical resistivity and earthquake generation in the crust of West Anatolia, Turkey. *Tectonophysics* **445**, 49–65 (2007).
- Bedrosian, P. A. *et al.* Geophysical images of the creeping segment of the San Andreas Fault: implications for the role of crustal fluids in the earthquake process. *Tectonophysics* **385**, 137–158 (2004).
- Tenthorey, E. & Stephen, F. C. Reaction-enhanced permeability during serpentinite dehydration. *Geology* **31**, 921–924 (2003).
- Ozacar, A. A. & Zandt, G. Crustal structure and seismic anisotropy near the San Andreas Fault at Parkfield, California. *Geophys. J. Int.* **178**, 1098–1104 (2009).
- Becken, M. *et al.* A deep crustal fluid channel into the San Andreas Fault system near Parkfield, California. *Geophys. J. Int.* **173**, 718–732 (2008).
- Wiersberg, T. & Erzinger, J. A helium isotope cross-section study through the San Andreas Fault at seismogenic depths. *Geochem. Geophys. Geosyst.* **8**, Q01002 (2007).
- Zoback, M., Hickman, S. & Ellsworth, W. Scientific drilling into the San Andreas fault zone. *Eos Trans. AGU* **91**, 197–199 (2010).
- Rodi, W. & Mackie, R. L. Nonlinear conjugate gradients algorithm for 2D magnetotelluric inversion. *Geophysics* **66**, 174–187 (2001).
- Becken, M. & Burkhardt, H. An ellipticity criterion in magnetotelluric tensor analysis. *Geophys. J. Int.* **159**, 69–82 (2004).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This research was supported by the German Science Foundation (DFG) and the Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences. Instruments were provided by the Geophysical Instrument Pool Potsdam and the instrument facility for Electromagnetic Studies of the Continents. Seismicity data are taken from the catalogue provided by the Northern California Earthquake Data Center, Northern California Seismic Network, US Geological Survey, Menlo Park (<http://www.ncedc.org>). Fault zone locations are from the California Geological Survey, Geologic Data Map Series No. 6. H. Zhang and D. Shelly provided their tremor source location estimates. This work would not have been possible without field support and the cooperation of numerous landowners.

Author Contributions All authors participated in the various field experiments. M.B. processed the time series, analysed and inverted the magnetotelluric data, and validated the resistivity models. All authors contributed to the interpretation and the manuscript. O.R. was principal investigator of the DFG/GFZ-funded research proposals.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.B. (michael.becken@uni-muenster.de).

METHODS

Five-channel magnetotelluric recordings were made at 250 sites during three surveys in the years 2005, 2007 and 2008. All sites were equipped with induction-coil magnetometers and non-polarizable electrodes (spanning orthogonal electric dipoles of length ~ 60 m) to measure the three components of the magnetic field and the horizontal electric field components. Three to four days of recording with this configuration yielded sufficient data to cover a broadband period range of magnetotelluric field variations, from 0.01 s to more than 1,000 s. In addition, 120 of the sites were equipped with fluxgate magnetometers and operated for two to five weeks to cover the long-period magnetotelluric period range up to 20,000 s. Up to 45 simultaneously recording instruments were deployed, yielding great flexibility in subsequent remote and multi-remote reference processing of the data. In addition to our own measurements, we included some of the earlier magnetotelluric data coincident with the central parts of profile 2 near SAFOD and profile 5 across the Cholame valley^{31,32}.

The primary step in magnetotelluric data analysis is to estimate the linear transfer functions between the horizontal electric and magnetic field components (impedance tensor, Z) and between the vertical and horizontal magnetic field components (vertical magnetic transfer function, T). Magnetotelluric transfer functions are diagnostic of subsurface electrical resistivity structure and are thus the quantities used in the subsequent construction of a numerical resistivity model. Standard robust processing techniques in the frequency domain^{33–35} produced high-quality transfer functions for periods longer than 5 s (Supplementary Fig. 1), but often failed to separate the shorter-period magnetotelluric signal from strong periodic noise with a fundamental period of ~ 5.1 s that was correlated over all channels and over wide areas. Noise levels were particularly high near buried gas pipelines, suggesting that cathodic protection currents commonly imposed on pipelines were the source of this noise. Remote reference techniques³³, multi-remote reference³⁶ and rejection of individual stacks based on phase constraints³⁵ could not substantially improve the short-period data quality. Instead, we preprocessed the time series with a fractional delay filter that proved to be a powerful tool in removing harmonic noise. The filter subtracts a delayed version of the time series from itself, which causes destructive interference at the delay period and all harmonics. The delay period corresponds to the fundamental period of the harmonic noise signal. Because the fundamental noise period does not fit into our sampling schemes, the fractional delay filter is designed to perform an interpolation of the signal to compute the delayed signal. The interpolation can be made exact because the signal is band limited³⁷. For subsequent remote reference processing, we applied the same delay filter also to the non-noisy reference sites. With this processing, we could notably improve the data quality at short periods (Supplementary Fig. 1).

We used two-dimensional (2D) inversion techniques to convert the estimated magnetotelluric transfer functions into the resistivity models discussed in the main text. Magnetotelluric transfer functions inherently contain measures that can justify a 2D interpretation and that define the geoelectric strike direction to which the data must be rotated before model computation. We extracted these measures from the impedance data by decomposition methods that attempt to separate the regional 2D magnetotelluric response from the ‘galvanic distortion effect’ of charge accumulations at near-surface resistivity heterogeneities³⁸. We used the impedance tensor decomposition technique described in ref. 30 to analyse the distortion effect and the dimensionality of the data and, because we found conditions that generally justified a 2D interpretation, to estimate the rotation angle corresponding to the geoelectric strike direction. Magnetotelluric strike estimates, however, have an inherent ambiguity of 90° . Considering the orientations of the major structural trends in the area (the coastline, the low-resistivity sedimentary fill of the San Joaquin valley, the SAF and associated geological units within the Coast Ranges), which are subparallel features elongated in the NW–SE direction, we constrained the strike estimates to the range 0° – 90° W. We determined consistent multi-period strike estimates for each magnetotelluric site separately, ranging between 35° W and 60° W, and a global strike estimate of 41° W using all available sites and periods jointly (Fig. 1a, inset, and main text). This geoelectric strike direction is in good agreement with dominant geological trends, particularly with the strike of the SAF in the segment under investigation. Nevertheless, variations in impedance phases oblique to the strike are also visible, and indicate, in conjunction with the off-profile orientation of induction vectors, the presence of three-dimensional (3D) effects in some regions (see Supplementary Figure 1).

In rotated coordinates, and in a true 2D environment with the x axis oriented along-strike, the Z_{xy} and Z_{yx} components of the impedance tensor then correspond to the E - and B -polarization impedances associated with along-strike and across-strike currents, respectively. Along-strike current concentrations are also associated with a vertical magnetic field anomaly at the surface. In constructing our 2D model, we included the off-diagonal Z_{xy} and Z_{yx} components of the

impedance tensor rotated by -41° (that is, 41° anticlockwise) and the projection of the vertical magnetic transfer function onto the across-strike y axis (T_y in rotated coordinates).

Because galvanic distortion has an irresolvable scaling effect on the apparent resistivities³⁹, we concentrated on fitting the impedance phases and the vertical magnetic transfer functions during the inversion. In practice, we down-weighted only the E -polarization apparent resistivity because 2D modelling can mimic the scaling effect for the B polarization. Down-weighting of data subsets is achieved in practice by increasing their error bars, which are used to weight the data residuals during the inversion. Here we imposed a global error floor of 100% on the E -polarization apparent resistivities, whereas lower error floors of 10% were imposed on the B -polarization apparent resistivities. For the impedances phases and for T_y , we used absolute error floors of 1.5° and 0.03, respectively.

Two-dimensional inversion of data with 3D effects may result in biased models or in models that contain phantom structures. Because the E -polarization impedance is considered most susceptible to 3D effects⁴⁰, some authors prefer to invert only B -polarization data or a combination of B -polarization apparent resistivities and phases and vertical magnetic transfer functions^{13,41,42}. This approach implicitly corresponds to a rigorous down-weighting of the E -polarization apparent resistivity and phase data with infinite error floors. However, accumulations of charges due to across-strike current flow within a laterally heterogeneous crust may dominate the B -polarization response even at long periods and mask the response of anomalies within the deeper crust and upper mantle⁴³. Therefore, valuable information about the deeper resistivity structure contained in the E -polarization responses could be discarded with this rigorous approach. A heuristic approach to reducing the influence of 3D effects on the inversion result without rejecting these data completely is to down-weight data subsets relative to their departure from pure 2D conditions. Here we used the ellipticity of the impedance tensor columns to measure the deviation of the data from pure 2D conditions in the presumed strike coordinate system, and adjusted the data errors such that the ellipticities and, hence, the departure from 2D conditions, vanish to within the modified errors²⁶. We applied this technique to the impedance data at every period and site individually to obtain modified error bars for the entire data set. The error bars of the vertical magnetic transfer function (T_y) were modified accordingly; here we used the absolute value of the off-profile component (T_x) as the error bar on T_y . These modified error bars were used as floor errors where they exceeded the global error floor defined above.

Inverse models were computed with a widely used nonlinear conjugate gradient 2D inversion program²⁹. The inversion seeks to minimize a Tikhonov-regularized functional, $\phi = \chi^2 + \tau L[\log(\rho)]$, as a trade-off between a structure penalty functional, $\tau L[\log(\rho)]$, and the sum of the squared data residuals weighted by their variances (χ^2). The parameter τ quantifies the trade-off and $L[\log(\rho)]$ is a weighted integral of the squared Laplacian of the logarithm of the model resistivity, ρ .

For the computation of χ^2 , lower thresholds for the data errors have been defined in terms of error floors and in terms of errors measuring the departure from the 2D assumption (see above). Statistically estimated data errors (produced with the robust time series processing code) enter the computation of χ^2 only where they exceed the former error measures.

The Laplacian is computed in the model domain rather than in the space domain, and is thus dependent on the model grid. Effectively, increasing horizontal cell dimensions with distance from the profile increases horizontal smoothing, whereas increasing vertical cell dimension with depth results in increased vertical smoothing. Therefore, we applied weight factors, α and β , to control the ratio of vertical to horizontal smoothing and to increase horizontal smoothing with increasing depth, respectively. Furthermore, the inversion program makes it possible to apply these weight factors to model domains containing cells with dimensions exceeding definable floor sizes in the horizontal and the vertical directions. For instance, defining the floor size in the vertical direction to correspond to the height of the cells at some depth, z_w , means that the weight factors α and β become effective only at depths $z > z_w$. We used $\alpha = 1$ throughout and tested β values between 1 and 3. We found that models computed with $\beta = 1$ over the entire model domain and models computed with $\beta = 3$ over only the model domain deeper than ~ 20 km contained very similar structures in the upper ~ 40 km, but that the former seemed vertically stretched at greater depths. Therefore, we used the latter parameter combination ($\alpha = 1$, $\beta = 3$).

The trade-off parameter, τ , which controls the relative contribution of data residuals and the structure penalty to the minimized objective function, was determined from the knee of an L-curve plot, that is, a cross-plot of data residuals and the model norm, here computed for τ values between 1 and 1,000. The optimal trade-off parameter was found to be close to $\tau = 30$, a value that is typical for the settings we used. We verified our inversion settings by comparing our inversion models with those produced with the OCCAM program⁴⁴, which inverts simulta-

neously for resistivity and the regularization parameter. The inversion programs produced similar results (not shown).

For the inversion, the Pacific Ocean was included as a fixed, a-priori structure in the starting model. Furthermore, we constrained the parts of the model deeper than 150 km (where the sensitivity of our data is poor and lateral resolution is limited by the length of our profiles) to a resistivity of 50 Ω m to avoid artificial structures appearing at the lower model boundary. Inversion tests confirm that the resistivity of the basal half-space has no strong influence on any of the major features of the inversion models discussed (at depths <60 km). For the final inversion models shown in this paper, we used these model constraints and the same inversion settings for all seven profiles; only the model grid was adjusted to match varying site distributions along the profiles. In Supplementary Figs 2–8, we show the observed data, the data predicted by the final inversion models and the data residuals in terms of distance–period pseudosections.

We tested the robustness of structural details contained in the final inversion models with 2D forward modelling and constraint inversion strategies (Supplementary Figs 3, 6 and 7), and with 3D forward modelling (Supplementary Fig. 9). Details are provided in Supplementary Information.

31. Unsworth, M. J., Malin, P., Egbert, G. D. & Booker, J. R. Internal structure of the San Andreas Fault at Parkfield. *Calif. Geol.* **25**, 359–362 (1997).
32. Unsworth, M. J., Bedrosian, P. A., Eisel, M., Egbert, G. D. & Siripunvaraporn, W. Along strike variations in the electrical structure of the San Andreas fault at Parkfield. *Geophys. Res. Lett.* **27**, 3021–3024 (2000).
33. Egbert, G. D. & Booker, J. R. Robust estimation of geomagnetic transfer functions. *Geophys. J.* **87**, 173–194 (1986).
34. Ritter, O., Junge, A. & Dawes, G. J. K. New equipment and processing for magnetotelluric remote reference observations. *Geophys. J. Int.* **132**, 535–548 (1998).
35. Weckmann, U., Magunia, A. & Ritter, O. Effective noise separation for magnetotelluric single site data processing using a frequency domain selection scheme. *Geophys. J. Int.* **161**, 635–652 (2005).
36. Egbert, G. D. Robust multiple-station magnetotelluric data processing. *Geophys. J. Int.* **130**, 475–496 (1997).
37. Laakso, T. I., Välimäki, V., Karjalainen, M. & Laine, U. K. Splitting the unit delay—tools for fractional delay filter design. *IEEE Signal Process. Mag.* **13**, 30–60 (1996).
38. Groom, R. W. & Bailey, R. C. Decomposition of magnetotelluric impedance tensors in the presence of local three-dimensional galvanic distortion. *J. Geophys. Res.* **94**, 1913–1925 (1989).
39. Smith, J. T. Understanding telluric distortion matrices. *Geophys. J. Int.* **122**, 219–226 (1995).
40. Wannamaker, P. E. in *Three-Dimensional Electromagnetics* (eds Oristaglio, M. & Spies, B.) 349–374 (SEG, 1999).
41. Wannamaker, P. E. *et al.* Fluid generation and pathways beneath an active compressional orogen, the New Zealand Southern Alps, inferred from magnetotellurics data. *J. Geophys. Res.* **107**, 2117 (2002).
42. Booker, J. R., Favetto, A. & Pomposiello, M. C. Low electrical resistivity associated with plunging of the Nazca flat slab beneath Argentina. *Nature* **429**, 399–403 (2004).
43. Becken, M., Ritter, O. & Burkhardt, H. Mode separation of magnetotelluric responses in three-dimensional environments. *Geophys. J. Int.* **172**, 67–86 (2008a).
44. deGroot-Hedlin, C. & Constable, S. Occam's inversion to generate smooth two-dimensional models from magnetotelluric data. *Geophysics* **55**, 1613–1624 (1990).

Energetics and the evolution of human brain size

Ana Navarrete¹, Carel P. van Schaik¹ & Karin Isler¹

The human brain stands out among mammals by being unusually large. The expensive-tissue hypothesis¹ explains its evolution by proposing a trade-off between the size of the brain and that of the digestive tract, which is smaller than expected for a primate of our body size. Although this hypothesis is widely accepted, empirical support so far has been equivocal. Here we test it in a sample of 100 mammalian species, including 23 primates, by analysing brain size and organ mass data. We found that, controlling for fat-free body mass, brain size is not negatively correlated with the mass of the digestive tract or any other expensive organ, thus refuting the expensive-tissue hypothesis. Nonetheless, consistent with the existence of energy trade-offs with brain size, we find that the size of brains and adipose depots are negatively correlated in mammals, indicating that encephalization and fat storage are compensatory strategies to buffer against starvation. However, these two strategies can be combined if fat storage does not unduly hamper locomotor efficiency. We propose that human encephalization was made possible by a combination of stabilization of energy inputs and a redirection of energy from locomotion, growth and reproduction.

Brains are energetically expensive². The human brain is about three times larger than that of our closest living relative, the chimpanzee, and thus requires much more energy. However, relative whole-body energy consumption rates of individuals at rest are about equal in the two species³, which raises the question of how humans manage to cover the energetic requirements of their much enlarged brains. One of the best-known attempts to solve this central riddle of human evolution is the expensive-tissue hypothesis, proposed by Aiello and Wheeler in 1995¹. It postulates an evolutionary trade-off (although obviously not an immediate physiological one) between the size of the brain and that of the digestive tract in anthropoid primates. Thus, if other processes have reduced a species' energetic needs of digestion, it should be able to evolve a relatively larger brain. It has therefore been suggested that early hominins evolved larger brains as a dietary shift towards more meat¹, cooked food and underground tubers⁴ gradually allowed for a smaller digestive tract.

The proposed trade-off would gain plausibility as a general principle if it were confirmed in other mammals. As stressed by Aiello *et al.*⁵, empirical support for a negative correlation across anthropoid primate species was weak from the beginning (see Supplementary Information, section 1 for a re-analysis), and subsequent comparative studies in other taxa remained ambiguous, with positive support in fish⁶, but not in bats or birds^{7,8}. Yet, this highly intuitive idea has found broad acceptance in palaeoanthropology⁹ and many other fields^{10–12}, and is fuelling public discussions about the optimal human diet. Thus, a proper empirical test of the expensive-tissue hypothesis across a broad array of taxa is urgently needed, but has not been conducted—until now—owing to a lack of morphological data, nor has there been an examination of the broader trade-offs among other expensive organs predicted by an extension of this hypothesis⁷.

Here we examine the presence of correlated evolution between the size of various visceral organs (heart, lungs, stomach, intestines, kidneys, spleen and liver) and that of the brain in a new sample of 100 mammalian species, including 23 primate species (see Supplementary

Data). In this analysis, it is crucial to control for body size, but the usual measure taken for this, body mass, is highly affected by variation in the size of adipose depots. This variation may confound or even reverse the direction of correlations among organs (Supplementary Fig. 2 and Supplementary Table 4b). Here, we therefore used fat-free body mass as the best proxy for body size.

Contrary to the predictions of the expensive-tissue hypothesis, we found no negative correlations between the relative size of the brain and the digestive tract, other expensive organs or their combined sum among mammals or within non-human primates, controlling for fat-free body mass, even though statistical power was sufficient to detect these negative correlations if they existed (see Table 1). We also did not find any trade-offs among other expensive organs (Fig. 1). These results therefore refute the expensive-tissue hypothesis as a general principle to explain the interspecific variation of relative brain size in mammals. In our view, this finding reduces the plausibility of the argument that human encephalization was made possible by a reduction of the digestive tract^{1,5}.

Energy trade-offs with other tissues that are less expensive but very abundant⁷ may nonetheless explain part of brain size variation. For instance, adipose depots make up an appreciable proportion of body mass in some mammals¹³. Although not metabolically expensive, adipose tissue has an energetic cost because it has to be carried around and may increase predation-induced mortality (see Supplementary Information 3.7). Fat stores enable animals to cope with periods of reduced food intake and thus act as a physiological buffer against starvation. On the other hand, relatively large brains have also been proposed to act as cognitive buffers against starvation^{14,15}. It is therefore possible that encephalization and fat storage are complementary strategies to buffer against starvation. In our sample of mammals, there is indeed a negative correlation between brain size and the size of fat stores, controlling for fat-free body mass (Table 1 and Supplementary Information 3.1), with the exception of primates (but see Supplementary Information 3.6). This negative relationship becomes stronger if potential error variation is removed, for instance by analysing only wild-caught females (Fig. 2 and Supplementary Information 3.4). The strongest trade-off between fat storage and brain size evolution is expected in taxa that exhibit a high cost of transport for increased whole body mass, such as climbing or flying mammals

Table 1 | Regressions of brain volume on other organ masses

Organ	Mammals (N = 100)		Primates (N = 23)	
	β	P value	β	P value
Heart	0.15	0.13	0.65	0.007
Lungs	−0.03	0.73	0.44	0.07
Kidneys	0.01	0.92	0.34	0.08
Liver	−0.02	0.84	0.20	0.30
Digestive tract	0.16	0.06	0.48	0.005
Stomach	0.15	0.042	0.17	0.19
Intestines	0.11	0.15	0.41	0.008
Spleen	−0.02	0.60	0.15	0.13
Visceral organs	0.05	0.64	0.50	0.029
Adipose depots	−0.07	0.017	−0.01	0.92

Data were controlled for fat-free body mass. Statistical details and the results of the N = 45 species subsample are listed in Supplementary Information 3.1.

¹Anthropological Institute and Museum, University of Zurich, Winterthurerstrasse 190, 8057 Zurich, Switzerland.

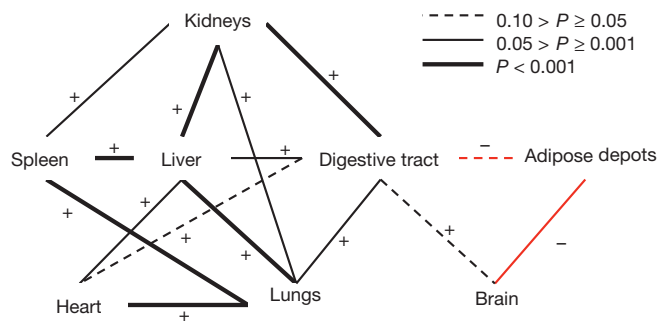


Figure 1 | Correlations between the masses of visceral organs, brains and adipose depots in mammals. The analysis is based on a sample of 100 mammalian species and controls for phylogenetic relationships and fat-free body mass. Statistical details are listed in the Supplementary Information 3.1.

and birds. The only animals that can easily combine both strategies of fat storage and brain enlargement may be those that do not face an increased cost of transport for increased whole body mass, for example, aquatic mammals or large bipeds¹⁶. However, more detailed studies of seasonal variation in body mass are needed to investigate which conditions or lifestyles favour one or the other, or a combination of both strategies.

Refuting the expensive-tissue hypothesis raises questions about the determinants of the evolution of the greatly enlarged human brain. Although there are various cognitive benefits to increased brain size¹⁷, empirical evidence shows that a focus on the energy costs of growing and maintaining brain tissue helps to explain the interspecific variation in brain size¹⁸. This approach has recently been synthesized in a general energy-based framework¹⁹, which incorporates earlier ideas on energetic aspects of brain size evolution^{1,5,18}. Figure 3 depicts the two possible pathways enabling increased encephalization from a given ancestral state: additional or stabilized energy inputs, and redirection of energy from other functions. Here we apply this framework to develop hypotheses for the remarkable increase of brain size during the evolution of the genus *Homo*.

Larger brains are sometimes paid for by a permanent increase in net energy intake of an organism, as indexed by its basal metabolic rate (BMR); this is shown by the positive correlation between BMR and brain size in a large sample of placental²⁰ and marsupial mammals²¹. This was confirmed in the present data set, in which we could control

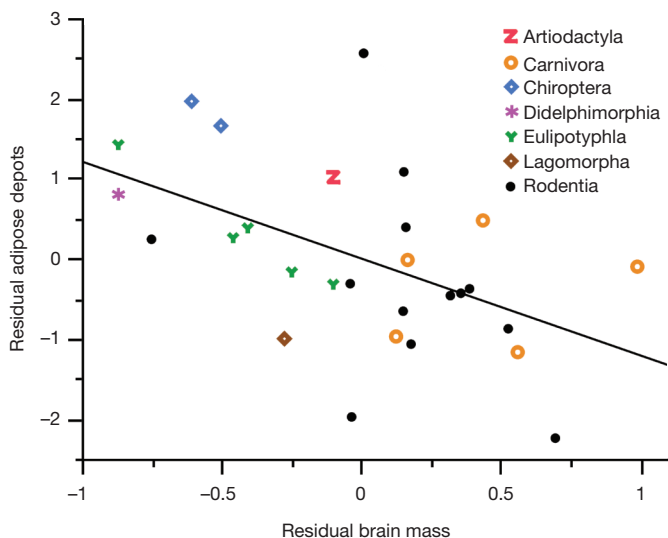


Figure 2 | Correlation between residual brain mass and residual adipose depots mass in wild-caught female mammals, controlling for fat-free body mass. Species-level values: $N = 28$ species, $r^2 = 0.258$, $P = 0.006$. PGLS: $\lambda = 1.00$, $\beta = -0.12$, t value = -3.42 , $P = 0.002$.

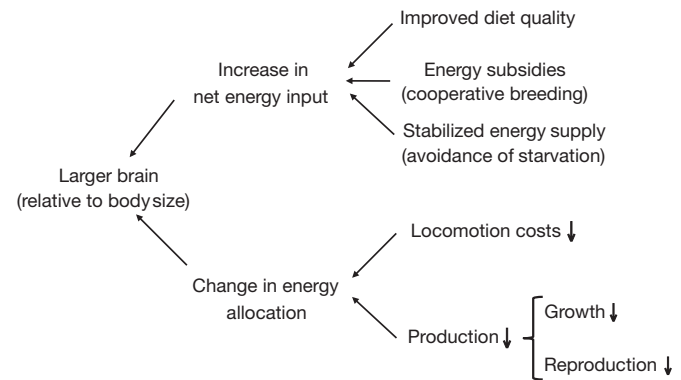


Figure 3 | The expensive-brain framework¹⁹ proposes complementary pathways for an adaptive increase in relative brain size. First, brains can get larger when energy inputs are stabilized on a higher level (higher total metabolic turnover²⁰) through an increase in mean dietary quality (for example, more animal fat and protein in early *Homo*^{4,22,24}), energy subsidies from other individuals (for example, cooperative breeding, allomaternal care^{19,21}) or by reducing fluctuations in energy inputs (for example, cognitive solutions¹⁵, including culture). Second, at constant total energy intake, energy allocation to other functions may be reduced, such as locomotion (for example, efficient bipedalism^{27,28}) or production (for example, slower life history pace³⁰).

for fat-free body mass ($N = 64$, PGLS, brain size as response, fat-free body mass and body mass associated with BMR measurements as covariates, effect of BMR: $\lambda = 0.96$, $P = 0.026$, $\beta = 0.24$). We humans exhibit the BMR expected for a mammal or primate of our body mass, but because we have much larger adipose depots (about 14–26% in healthy adults²²) than chimpanzees and bonobos (about 3–10% (ref. 23)), human BMR relative to fat-free body mass is appreciably higher than theirs²⁴. Therefore, if extant apes are representative of the last common ancestor, brain enlargement during human evolution was partially paid for through a permanent increase in net energy intake.

Starting with Early Pleistocene *Homo*, this increase could have come from any of the three sources listed in Fig. 3. First, they improved diet quality as indicated by increased consumption of meat and bone marrow¹ and by tool-assisted food processing, at one point including cooking⁴. Second, despite having moved into highly seasonal habitats⁹ they reduced temporal fluctuations in energy budgets by cognitive buffering²⁵, which is also known for other primates¹⁵ and birds¹⁴. Third, provisioning and food sharing probably arose with the adoption of cooperative breeding and substantial meat acquisition among the earlier representatives of the genus *Homo*^{4,26}. Comparative research suggests that such energy subsidies for reproducing females and dependent offspring can support increased brain size^{19,21}.

The second pathway to brain enlargement is increased energy allocation to the brain by savings on other expensive functions, although the expensive-tissue hypothesis for organs is no longer supported. One likely trade-off could be found between brain size and the costs of locomotion. The efficient form of bipedal locomotion that arose with the transition from australopithecines to early *Homo*²⁷ could have led to major reductions in energy expenditure in two ways. On one hand, its low costs in comparison with the climbing and quadrupedal locomotion of nonhuman apes²⁸ should have lowered daily energy expenditure on locomotion⁷, and on the other hand, bipedalism may reduce the effect of increased weight due to adipose depots on the energy costs of locomotion (Supplementary Information 3.7). A second potential trade-off would be the one between brain size and production, comprising both growth and reproductive effort, which has been demonstrated for mammals^{19,29}. Beginning with early *Homo* our lineage has increased brain size and reduced the pace of life history³⁰, but nonetheless increased birth rates due to cooperative breeding.

In sum, we do not claim unique processes operating exclusively in human evolution. All these processes are known to operate among mammals in general. We propose that during human evolution improved diet quality, allomaternal subsidies, cognitive buffering, reduced locomotion costs and reduced allocation to production all operated simultaneously, thus enabling the extraordinary brain enlargement in our lineage.

METHODS SUMMARY

Following a strict protocol (Supplementary Information 2), A.N. dissected a large number of mammalian specimens obtained from various sources. Visceral organs (kidneys, spleen, liver, stomach, intestines, heart and lungs) were separated, cleaned, emptied, and immediately weighed. As skulls had to be preserved intact, cranial capacity was determined and converted into an estimate of brain mass. We excluded individuals that were immature, emaciated, pregnant, or exhibited visible organ pathologies from analysis. Hence, our final sample included 191 specimens from 100 species, with a bias towards carnivores and primates (Supplementary Data and Supplementary Information 3.2). The sample size of 100 species yields a power of 0.8 for these analyses, which was determined a priori using a published data set of 39 mammal species.

Traditionally, whole body mass has been used for controlling for body size effects in comparative analyses. However, this measure is highly affected by variation in the size of adipose depots. Even if two species have a similar fat-free body mass and body composition, large adipose depots in one species result in organs that seem relatively small in comparison to those of a species with smaller adipose depots. Therefore, correlations between organs are expected to be mostly positive if we controlled for whole body mass (Supplementary Fig. 2). To avoid this bias, we used fat-free body mass to control for body size effects. For 45 species in our sample, fat-free body mass was known because adipose depots of whole body were measured directly. For the other 55 species, a proxy was calculated from the measurement of abdominal adipose depots (Supplementary Information 3.1 and 3.3).

All analyses took phylogenetic relatedness into account (Supplementary Fig. 3). Additionally, we tested sex-specific and wild/captive subsamples separately (Supplementary Information 3.4), and investigated potential autocorrelation effects (Supplementary Information 3.5).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 23 May; accepted 12 October 2011.

Published online 9 November 2011.

1. Aiello, L. C. & Wheeler, P. The expensive-tissue hypothesis—the brain and the digestive system in human and primate evolution. *Curr. Anthropol.* **36**, 199–221 (1995).
2. Mink, J. W., Blumenshine, R. J. & Adams, D. B. Ratio of central nervous system to body metabolism in vertebrates—its constancy and functional basis. *Am. J. Physiol.* **241**, R203–R212 (1981).
3. Bruhn, J. M. & Benedict, F. G. The respiratory metabolism of the chimpanzee. *Proc. Am. Acad. Arts Sci.* **71**, 259–326 (1936).
4. Wrangham, R. *Catching Fire: How Cooking Made Us Human* (Basic Books, 2009).
5. Aiello, L. C., Bates, N. & Joffe, T. in *Evolutionary Anatomy of the Primate Cerebral Cortex* (eds Dean, F. & Gibson, K.) 57–78 (Cambridge Univ. Press, 2001).
6. Kaufman, J. A., Hladik, C. M. & Pasquet, P. On the expensive-tissue hypothesis: independent support from highly encephalized fish. *Curr. Anthropol.* **44**, 705–707 (2003).
7. Isler, K. & van Schaik, C. P. Costs of encephalisation: the energy trade-off hypothesis tested on birds. *J. Hum. Evol.* **51**, 228–243 (2006).
8. Jones, K. E. & MacLarnon, A. M. Affording larger brains: testing hypotheses of mammalian brain evolution on bats. *Am. Nat.* **164**, E20–E31 (2004).

9. Potts, R. Environmental hypotheses of hominin evolution. *Yearb. Phys. Anthropol.* **107**, 93–136 (1998).
10. Mau, M., Südekum, K.-H. & Kaiser, T. M. Why cattle feed much and humans think much—new approach to confirm the expensive tissue hypothesis by molecular data. *Biosci. Hypotheses* **2**, 205–208 (2009).
11. Pfeufferle, A. D. et al. Comparative expression analysis of the phosphocreatine circuit in extant primates: implications for human brain evolution. *J. Hum. Evol.* **60**, 205–212 (2011).
12. Santoro, S. et al. Preliminary results from digestive adaptation: a new surgical proposal for treating obesity, based on physiology and evolution. *Sao Paulo Med. J.* **124**, 192–197 (2006).
13. Pond, C. M. *The Fats of Life* (Cambridge Univ. Press, 1998).
14. Sol, D. Revisiting the cognitive buffer hypothesis for the evolution of large brains. *Biol. Lett.* **5**, 130–133 (2009).
15. van Woerden, J. T., van Schaik, C. P. & Isler, K. Effects of seasonality on brain size evolution: evidence from strepsirrhine primates. *Am. Nat.* **176**, 758–767 (2010).
16. Garland, T. Scaling the ecological cost of transport to body mass in terrestrial mammals. *Am. Nat.* **121**, 571–587 (1983).
17. Reader, S. M., Hager, Y. & Laland, K. N. The evolution of primate general and cultural intelligence. *Philos. Trans. R. Soc. B* **366**, 1017–1027 (2011).
18. Martin, R. D. Relative brain size and basal metabolic rate in terrestrial vertebrates. *Nature* **293**, 57–60 (1981).
19. Isler, K. & van Schaik, C. P. The expensive brain: a framework for explaining evolutionary changes in brain size. *J. Hum. Evol.* **57**, 392–400 (2009).
20. Isler, K. & van Schaik, C. P. Metabolic costs of brain size evolution. *Biol. Lett.* **2**, 557–560 (2006).
21. Isler, K. Energetic trade-offs between brain size and offspring production: marsupials confirm a general mammalian pattern. *Bioessays* **33**, 173–179 (2011).
22. Wells, J. C. K. *The Evolutionary Biology of Human Body Fatness* (Cambridge Univ. Press, 2009).
23. Zihlman, A. L. in *The Pygmy Chimpanzee* (ed. Susman, R. L.) 179–200 (Plenum Press, 1984).
24. Aiello, L. C. & Wells, J. C. K. Energetics and the evolution of the genus *Homo*. *Annu. Rev. Anthropol.* **31**, 323–338 (2002).
25. Kaplan, H., Hill, K., Lancaster, J. & Hurtado, A. M. A theory of human life history evolution: diet, intelligence, and longevity. *Evol. Anthropol.* **9**, 156–185 (2000).
26. Burkart, J. M., Hrdy, S. B. & van Schaik, C. P. Cooperative breeding and human cognitive evolution. *Evol. Anthropol.* **18**, 175–186 (2009).
27. Pontzer, H. et al. Locomotor anatomy and biomechanics of the Dmanisi hominins. *J. Hum. Evol.* **58**, 492–504 (2010).
28. Pontzer, H., Raichlen, D. A. & Sockol, M. D. The metabolic cost of walking in humans, chimpanzees, and early hominins. *J. Hum. Evol.* **56**, 43–54 (2009).
29. Isler, K. & van Schaik, C. P. Why are there so few smart mammals (but so many smart birds)? *Biol. Lett.* **5**, 125–129 (2009).
30. Dean, C. et al. Growth processes in teeth distinguish modern humans from *Homo erectus* and earlier hominins. *Nature* **414**, 628–631 (2001).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank R. D. Martin and J. Wermuth for sharing the Chivers data set, J. van Woerden for sharing her endocranial volume data, and M. Genoud for sharing his revised compilation of mammalian BMR values. Specimens were provided by numerous institutions, museums and colleagues (Supplementary Information 2). We acknowledge valuable comments by L. Aiello and R. D. Martin. Financial support was provided by the Swiss National Science Foundation (grant number 3100A0-117789), the A.H. Schultz-Stiftung and the European Integrated Activities grant SYNTHESIS (grant application number HU-TAF-4916).

Author Contributions K.I. and C.P.v.S. designed the project. A.N. performed the pilot study and collected the data. A.N. and K.I. performed the analyses and all three authors wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.N. (a.navarrete@aim.uzh.ch) or K.I. (kisler@aim.uzh.ch).

METHODS

Power analysis. To estimate the sample size needed to detect a correlation between brain size and digestive tract mass in mammals, controlling for fat-free body mass, we used an independent data set of 39 mammalian species³¹. In a multiple regression, digestive tract mass has a standard error of the residual error, σ , of 0.343, and a raw effect size δ of 0.096. Therefore, for a level of significance of $\alpha = 0.05$, the sample size required to achieve a power of 0.8 is 103 species. We thus aimed at collecting data from more than 100 mammal species.

Specimens. Four hundred and fifty four specimens of 133 mammal species were obtained from various sources and dissected following a strict protocol (Supplementary Information 2) by A.N. Visceral organs (kidneys, spleen, liver, stomach, intestines, heart and lungs) were separated, cleaned, emptied, and immediately weighed. As skulls had to be preserved intact, cranial capacity was determined using the seed filling method³², and converted into an estimate of brain mass by multiplying it by 1.036 (ref. 33). Specimens were excluded from analyses if they were juvenile or subadult, emaciated, pregnant, previously stored in formalin or alcohol, had visible pathologies of the organs (such as tumours or internal parasites), a broken neurocranium, unknown body mass before dissection, or if the organ measurements were incomplete. Our final sample included 191 specimens from 100 species, with a bias towards larger orders and especially carnivores and primates (see Supplementary Information 3.2 for species and family coverage and an additional analysis on subfamily level). Species values were obtained by calculating the average of male and female specimens (see Supplementary Information 3.4 for an analysis of sex-specific and wild/captive subsamples). Intestine mass was defined as the sum of ileum, caecum and colon mass, and digestive tract mass as the sum of stomach and intestines mass. BMR data of the species of our sample or of closely related taxa were taken from the literature. Data and sources are listed in Supplementary Data.

Adipose depots. For 45 species in our sample, adipose depots of the whole body were measured directly. For the other 55 species, only abdominal adipose depots were measured, from which we calculated a proxy of total adipose depots by scaling the abdominal depots mass with a factor 3.419. This scaling factor was derived from a comparison of the two measurements for 292 individuals for which body mass and one of the two adipose depots measurements was available (Supplementary Fig. 1). Alternatively, the total mass of adipose depots was calculated from abdominal mass using a prediction equation derived from nine specimens for which both measurements were available (results shown in Supplementary Information 3.3). Fat-free body mass was calculated as whole body mass minus total adipose depots mass.

Statistical analyses. All variables were log-transformed and phylogenetic regressions were run using `pglmEstLambda` in the CAIC³⁴ package in R³⁵. This function uses the PGLS method, estimating λ as an index of the amount of phylogenetic autocorrelation in the data. If λ is 0, species values are phylogenetically independent

and the analysis is equivalent to a species means least-squares regression. If λ is close to 1, the phylogenetic signal implies that trait evolution follows Brownian motion, and the analysis is equivalent to the classic method of calculating independent contrasts. If analyses yielded unstable estimates of λ due to the small sample size within orders (λ not significantly different from both 0 and 1), we additionally ran the analyses with λ set to 0 or 1. The models included brain mass as response, body size as covariate and organ mass as effect.

Visceral organs, the brain and adipose depots are part of the same body, and therefore autocorrelation effects could be suspected to influence our results. Two methods to remove these effects are reported in the Supplementary Information 3.5, and the results corroborate our findings. Analyses were done both using the total sample of 100 species and the subsample with total adipose depot mass of 45 species.

Whole body mass versus fat-free body mass. Traditionally, whole body mass has been used for controlling for body size effects in comparative analyses. However, this measure is highly affected by variation in the size of adipose depots and using it to control for body size may have an effect on the correlation between organs. Even if two species have a similar fat-free body mass and body composition, large adipose depots in one species result in organs that seem relatively small in comparison to those of a species with smaller adipose depots. Therefore, correlations between organs are mostly positive if we control for whole body mass. This bias is expected to disappear if fat-free body mass is used to control for body size effects (Supplementary Fig. 2).

These relationships were confirmed by our analyses. Brain–organ correlations were mostly positive, if whole body mass was included in the model (Supplementary Tables 4 and 5). Controlling for fat-free body mass instead of whole body mass reduced or eliminated this bias in most groups, with the notable exception of primates, in which positive correlations between organs persist, and brain size is not negatively related to adipose depots mass. We argue (Supplementary Information 3.6) that these discrepancies are due to a combination of error and peculiar captivity effects in foregut fermenting primates, and that primates would follow the general mammal trend if more complete data were available.

31. Pitts, G. & Bullard, T. in *Body Composition in Animals and Man* (ed. Reit, J. T.) 45–70 (National Academy of Science Pub No. 1598, 1968).
32. Isler, K. *et al.* Endocranial volumes of primate species: scaling analyses using a comprehensive and reliable data set. *J. Hum. Evol.* **55**, 967–978 (2008).
33. Rehkämper, G., Frahm, H. D. & Zilles, K. Quantitative development of brain and brain structures in birds (Galliformes and Passeriformes) compared to that in mammals (Insectivores and Primates). *Brain Behav. Evol.* **37**, 125–143 (1991).
34. Orme, D., Freckleton, R. P., Thomas, G., Petzoldt, T. & Fritz, S. CAIC: *Comparative Analyses Using Independent Contrasts* (<http://r-forge.r-project.org/projects/caic>) (2009).
35. R Development Core Team. R: a language and environment for statistical computing (R Foundation for Statistical Computing, 2010).

A SUMOylation-defective MITF germline mutation predisposes to melanoma and renal carcinoma

Corine Bertolotto^{1,2,3*}, Fabienne Lesueur^{4†*}, Sandy Giuliano^{1,2*}, Thomas Strub⁵, Mahaut de Lichy⁴, Karine Bille¹, Philippe Dessen⁶, Benoit d'Hayer⁴, Hamida Mohamdi^{7,8,9}, Audrey Remenieras^{4†}, Eve Maubec^{7,10}, Arnaud de la Fouchardière¹¹, Vincent Molinié¹², Pierre Vabres¹³, Stéphane Dalle¹⁴, Nicolas Poulalhon¹⁴, Tanguy Martin-Denavit¹⁴, Luc Thomas¹⁴, Pascale Andry-Benzaquen¹⁵, Nicolas Dupin¹⁵, Françoise Boitier¹⁵, Annick Rossi¹⁶, Jean-Luc Perrot¹⁷, Bruno Labeille¹⁷, Caroline Robert¹⁸, Bernard Escudier¹⁸, Olivier Caron¹⁸, Laurence Brugières¹⁹, Simon Saule²⁰, Betty Gardie²¹, Sophie Gad²¹, Stéphane Richard^{21,22}, Jérôme Couturier²³, Bin Tean Teh^{24,25}, Paola Ghiorzo²⁶, Lorenza Pastorino²⁶, Susana Puig²⁷, Celia Badenas²⁷, Hakan Olsson²⁸, Christian Ingvar²⁹, Etienne Rouleau³⁰, Rosette Lidereau³⁰, Philippe Bahadoran³, Philippe Viel³¹, Eve Corda^{7,9}, Hélène Blanché⁹, Diana Zelenika³², Pilar Galan³³, The French Familial Melanoma Study Group[†], Valérie Chaudru^{7,9,34}, Gilbert M. Lenoir^{4,35}, Mark Lathrop^{9,32}, Irwin Davidson⁵, Marie-Françoise Avril¹⁵, Florence Demenais^{7,8,9}, Robert Ballotti^{1,2,3*} & Brigitte Bressac-de Paillerets^{4,7*}

So far, no common environmental and/or phenotypic factor has been associated with melanoma and renal cell carcinoma (RCC). The known risk factors for melanoma include sun exposure, pigmentation and nevus phenotypes¹; risk factors associated with RCC include smoking, obesity and hypertension². A recent study of coexisting melanoma and RCC in the same patients supports a genetic predisposition underlying the association between these two cancers³. The microphthalmia-associated transcription factor (*MITF*) has been proposed to act as a melanoma oncogene⁴; it also stimulates the transcription of hypoxia inducible factor⁵ (*HIF1A*), the pathway of which is targeted by kidney cancer susceptibility genes⁶. We therefore proposed that *MITF* might have a role in conferring a genetic predisposition to co-occurring melanoma and RCC. Here we identify a germline missense substitution in *MITF* (Mi-E318K) that occurred at a significantly higher frequency in genetically enriched patients affected with melanoma, RCC or both cancers, when compared with controls. Overall, Mi-E318K carriers had a higher than fivefold increased risk of developing melanoma, RCC or both cancers. Codon 318 is located in a small-ubiquitin-like modifier (SUMO) consensus site (ΨKXE) and Mi-E318K severely impaired SUMOylation of MITF. Mi-E318K enhanced MITF protein binding to the *HIF1A* promoter and increased its transcriptional activity compared to wild-type MITF. Further, we observed a global increase in Mi-E318K-occupied loci. In an RCC cell line, gene expression profiling identified a Mi-E318K signature related to cell growth, proliferation and inflammation. Lastly, the mutant protein enhanced melanocytic and renal cell clonogenicity, migration and invasion,

consistent with a gain-of-function role in tumorigenesis. Our data provide insights into the link between SUMOylation, transcription and cancer.

MITF encodes a member of the Myc supergene family of basic helix-loop-helix zipper transcription factors. It has a complex intron-exon structure, yielding protein products with different amino termini⁷. The M-isoform (expressed in the melanocyte lineage) of MITF regulates expression of a large set of genes promoting proliferation and invasion. For instance, it controls the expression of *MET*⁸ and *CDKN2A/p16INK4A*⁵, which have key roles in melanoma development. The role of the A-isoform of MITF (expressed in kidney) in renal cell transformation has yet to be elucidated. However, *MITF* belongs to the same family of transcription factors (the MiT family) as the *TFEC*, *TFEB* and *TFEB* genes; the two latter were identified as targets of somatic recurrent translocations in a RCC subtype found predominantly in children and young adults⁸.

We sequenced the *MITF* gene in 62 patients with melanoma and RCC. Five patients exhibited a germline heterozygous missense substitution, p.E318K (c.952G>A, in MITF isoform 4, NCBI accession NM_000248.3) (Fig. 1a). The frequency of this variant was significantly higher in patients with both melanoma and RCC (melanoma + RCC) than in 1,659 population-based cancer-free controls ($P = 1.3 \times 10^{-4}$). Carriers of the p.E318K variant (Mi-E318K) exhibited a 14-fold higher risk than controls for developing melanoma + RCC (Table 1). To investigate the effect of Mi-E318K on susceptibility to melanoma alone, we genotyped 603 affected patients that had undergone oncogenetic testing for melanoma susceptibility genes¹ and were negative for *CDKN2A* and *CDK4* mutations (Table 1). Mi-E318K

¹INSERM, U895 (équipe 1), Equipe labélisée Ligue Contre le Cancer, C3M, 06204 Nice, France. ²Université de Nice Sophia-Antipolis, UFR Médecine, 06204 Nice, France. ³Centre Hospitalier Universitaire de Nice, Service de Dermatologie, 06204 Nice, France. ⁴Service de Génétique, Institut de Cancérologie Gustave Roussy, 94805 Villejuif, France. ⁵Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS, INSERM, Université de Strasbourg, 67404 Illkirch, France. ⁶INSERM, UMR985, Institut de Cancérologie Gustave Roussy, 94805 Villejuif, France and Université Paris-Sud 11, 91405 Orsay. ⁷INSERM, U946, Genetic Variation and Human Diseases Unit, 75010 Paris, France. ⁸Université Paris Diderot, Sorbonne Paris Cité, Institut Universitaire d'Hématologie, 75010 Paris, France. ⁹Fondation Jean Dausset-Centre d'Etude du Polymorphisme Humain (CEPH), 75010 Paris, France. ¹⁰AP-HP, Hôpital Bichat, Service de Dermatologie, Faculté Paris Diderot, 75018 Paris, France. ¹¹Département de Pathologie, Centre Léon Bérard, 69008 Lyon, France. ¹²Department of Pathology, Hôpital Saint-Joseph, 75014 Paris, France. ¹³Department of Dermatology, Centre Hospitalier Universitaire, 21079 Dijon, France. ¹⁴Lyon 1 University and Centre Hospitalier Lyon Sud, Department of Dermatology, 69495 Pierre Bénite, France. ¹⁵AP-HP, Hôpital Cochin-Tarnier, Service de Dermatologie and Université Paris Descartes, 75006 Paris, France. ¹⁶Unité de Génétique Clinique, Hôpital Charles Nicolle, CHU Rouen, 76038 Rouen, France. ¹⁷Department of Dermatology, CHU Hôpital Nord, 42055 Saint-Etienne, France. ¹⁸Department of Medicine, Institut de Cancérologie Gustave Roussy, 94805 Villejuif, France. ¹⁹Department of Pediatrics, Institut de Cancérologie Gustave Roussy, 94805 Villejuif, France. ²⁰Institut Curie, CNRS UMR33 and INSERM U1021, Université Paris-Sud 11, 91405 Orsay, France. ²¹Génétique Oncologique EPHE-INSERM U753, Faculté de Médecine, Université Paris-Sud 11 and Institut de Cancérologie Gustave Roussy, 94805 Villejuif, France. ²²Centre Expert National Cancers Rares PREDIR, INCa/AP-HP, Service d'Urologie, Hôpital de Bicêtre, 94275 Le Kremlin-Bicêtre, France. ²³Service de Génétique, Institut Curie, 75005 Paris, France. ²⁴Laboratory of Cancer Genetics, Van Andel Research Institute, Grand Rapids, 49503 Michigan, USA. ²⁵NCCS-VARI Translational Research Laboratory, National Cancer Center Singapore, 16910 Singapore. ²⁶Department of Internal Medicine, University of Genoa, 16132 Genoa, Italy. ²⁷Dermatology Department and Biochemistry and Molecular Genetics, Melanoma Unit, Hospital Clinic, IDIBAPS, and CIBER de Enfermedades Raras, Instituto de Salud Carlos III, 08036 Barcelona, Spain. ²⁸Department of Oncology, Lund University and Hospital, Lund, 22185 Sweden. ²⁹Department of Surgery, Lund University and Hospital, Lund, 22185 Sweden. ³⁰Service d'Oncogénétique, Hôpital René Huguenin- Institut Curie, 92210 Saint-Cloud, France. ³¹Department of Pathology, Translational Research Laboratory and Biobank, Institut de Cancérologie Gustave Roussy, 94805 Villejuif, France. ³²Commissariat à l'Energie Atomique, Centre National de Génotypage, 91057 Evry, France. ³³INSERM, UMR557, INRA U1125, CNAM, Paris 13, CRNH IdF, 93000 Bobigny, France. ³⁴Université d'Evry Val d'Essonne, 91025 Evry, France. ³⁵Université Paris-Sud 11, Faculté de Médecine, 94275 Le Kremlin-Bicêtre, France. †Present addresses: Genetic Cancer Susceptibility group, International Agency for Research on Cancer, 69372 Lyon, France (F.L.); Genetic Department, Institut Paoli Calmettes, 13273 Marseille, France (A.R.).

*These authors contributed equally to this work.

†A list of authors and their affiliations appears at the end of the paper.

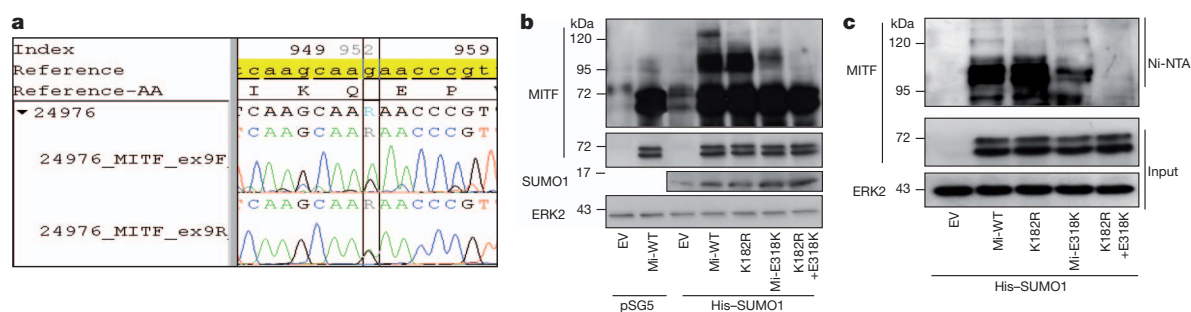


Figure 1 | Mi-E318K mutation impairs MITF SUMOylation. **a**, MITF sequence data of the c.952G>A, p.E318K germline substitution (Mi-E318K). **b**, HEK293 cells were co-transfected with empty pcDNA3 vector (EV) or pcDNA3 encoding wild-type (Mi-WT) or mutant (K182R, Mi-E318K or K182R+E318K)

frequency was significantly higher in these patients than in controls ($P = 7.8 \times 10^{-5}$). The carriers of Mi-E318K in this group exhibited >4-fold higher melanoma risk than controls (Table 1). In addition, Mi-E318K co-segregated with melanoma in three melanoma-prone families that had DNA available in relatives of index cases carrying Mi-E318K (Supplementary Fig. 1a). We also found that Mi-E318K was more strongly associated with multiple occurrences of primary melanomas than with single occurrence, irrespective of family history ($P = 0.02$, Supplementary Table 1). To investigate the effect of Mi-E318K in RCC susceptibility, we genotyped 164 patients with RCC referred to oncogenetic clinics, who were enriched for rare histological subtypes and who were wild type for the known RCC-predisposing genes⁶. These patients also exhibited a higher frequency of Mi-E318K than controls ($P = 0.008$) (Table 1). After correcting for multiple testing, the association between Mi-E318K and melanoma + RCC, melanoma only or RCC only remained significant, although marginally so for the RCC-only group ($P = 0.02$). The three groups showed similar Mi-E318K allele frequencies ($P = 0.10$). When all 829 patients with melanoma and/or RCC were pooled, Mi-E318K frequency was significantly different from controls ($P = 1.2 \times 10^{-6}$). The pooled group of Mi-E318K carriers had a greater than fivefold risk of developing melanoma, RCC or both cancers, as compared to controls (Table 1). The clinical and pathological features of the 27 Mi-E318K carriers are described in Supplementary Table 2. Principal component analysis (PCA) of single nucleotide polymorphism (SNP) data across the genome in more than 75% of cases and controls showed appropriate clustering of cases and controls after exclusion of a few subjects of non-European ancestry (Supplementary Fig. 1b). Association analysis of melanoma and/or RCC with Mi-E318K provided similar results when performed with or without adjusting for principal components (Supplementary Table 3). To assess whether Mi-E318K predisposed to the coexistence of melanoma and a second primary neoplasm other than RCC, we gathered data from European countries, but these data were insufficient for separate analyses of each type of cancer. Overall, we concluded from our findings that Mi-E318K is a rare substitution that confers an increased risk for developing melanoma and/or RCC. However, given the selection of patients who were part of case series undergoing genetic testing, further investigation of Mi-E318K in large series of unselected sporadic renal cancer and melanoma cases is merited.

Myc-tagged MITF and either with an empty vector or His-SUMO1 pSG5 vectors. Western blot analysis with antibodies to MITF, SUMO1 and ERK2 used as a loading control. **c**, Western blot analysis with antibodies to MITF and ERK2 of protein extracts before (Input) and after Ni-NTA affinity purification.

Previous studies have shown that MITF is expressed in a majority of melanomas and in a subset of kidney tumours⁹. Accordingly, our immunohistochemical analyses showed that MITF was expressed in both melanomas from patients carrying the germline mutation Mi-E318K ($n = 8$) and in melanomas from wild-type patients ($n = 8$) (Supplementary Fig. 2a, b). In addition, two out of seven RCCs from Mi-E318K carriers were positive for MITF (Supplementary Fig. 2c, d), suggesting that Mi-E318K might have a role in renal transformation. None of the six wild-type MITF RCC samples analysed in this study showed MITF labelling.

Mi-E318K occurs at a conserved position in MITF within a consensus motif (IKQE) that matches perfectly with the consensus sequence ΨKXE for covalent binding of SUMO¹⁰ (Supplementary Fig. 3a, b). Because SUMOylation is critically dependent on the acidic residue at +2 (E) of the acceptor lysine (K), we tested whether Mi-E318K affects MITF SUMOylation. Co-expression of histidine-tagged (His)-SUMO1 with wild-type MITF and western blot analysis revealed a 120-kDa band and a doublet above 95 kDa, in addition to the 55–65 kDa native doublet (Fig. 1b). The p.K182R mutation introduced within the second MITF SUMOylation site led to complete disappearance of the 120-kDa MITF form, but had little effect on the 95-kDa doublet. When glutamic acid residue 318 was replaced by a lysine, mimicking the germline mutation, we observed a strong decrease in the levels of all the SUMO-modified forms of MITF. Lastly, no SUMOylated forms were seen with the double-mutant [p.K182R;p.E318K]. Similar observations were made with co-expression of haemagglutinin-tagged (HA)-SUMO2 (Supplementary Fig. 4a), demonstrating that Mi-E318K affected both SUMO1 and SUMO2 modifications. Western blot analysis with anti-MITF antibody, after affinity purification of His-SUMO1-containing proteins on Ni-NTA columns, confirmed that the high molecular weight bands were indeed SUMO-modified MITF proteins and that Mi-E318K markedly reduced all the MITF SUMOylated forms (Fig. 1c). Altogether, these results demonstrated that codon 316 was the major SUMO acceptor site in MITF and that Mi-E318K severely impaired SUMO conjugation to MITF.

SUMOylation has been shown to orchestrate a variety of cellular processes, in part through the control of nucleo-cytoplasmic signal transduction. However, both Mi-E318K and wild type were detected

Table 1 | Frequency of the germline Mi-E318K substitution

Subjects	Number of non-carriers	Number of carriers‡	Total	Frequency of p.E318K	FET <i>P</i> value†	OR (95%CI)*
Controls	1,649	10	1,659	0.003	-	Reference
Patients with melanoma and/or RCC	802	27	829§	0.016	1.2×10^{-6}	5.55 (2.59–12.91)
Melanoma + RCC	57	5	62	0.040	1.3×10^{-4}	14.46 (3.74–48.04)
Melanoma only	586	17	603	0.014	7.8×10^{-5}	4.78 (2.05–11.75)
RCC only	159	5	164	0.015	0.008	5.19 (1.37–16.87)

*OR (95% CI) is the odds ratio (with 95% confidence interval) associated with the Mi-E318K carrier status.

†FET, Fisher's exact test for the difference in Mi-E318K allele frequency between each group of patients and controls.

‡All carriers are heterozygotes for the Mi-E318K substitution. Clinical and pathological features of the 27 carriers are described in Supplementary Table 2.

§A full description of the 829 patients is given in Methods.

||The Fisher's exact test showed no significant differences ($P = 0.10$) in allele frequency among the three groups of patients (melanoma + RCC, melanoma only, RCC only).

by immunofluorescence in the nucleus of transfected melanoma cells. This indicated that Mi-E318K did not affect MITF nuclear localization (Supplementary Fig. 4b, c).

Previous reports indicated that SUMOylation of MITF repressed its transcriptional activity, particularly when the targeted promoter contained multiple E-boxes¹⁰. Indeed, on a synthetic reporter gene comprising three E-boxes, Mi-E318K showed two- to threefold greater transcriptional activity than that of wild type (Fig. 2a). The double MITF mutant, [p.E318K;p.K182R], was more effective than Mi-E318K, and the single p.K182R mutant exhibited activity comparable to that of wild type. We next tested whether Mi-E318K affected MITF activity on physiological target promoters, such as *MET* and *HIF1A*, two MITF-regulated genes involved in cell survival and in melanocyte and kidney tumorigenesis. We also tested *CDKN2A* and *TYR* promoters, which are involved in melanocyte proliferation and differentiation processes, respectively. Wild type and mutant Mi-E318K showed similar activity on the *MET* (Fig. 2b), *TYR* and *CDKN2A* promoters (Supplementary Fig. 5), but Mi-E318K more efficiently activated the *HIF1A* promoter than wild type (Fig. 2c). This observation suggests that Mi-E318K alters MITF transcriptional activity on a subset of its target genes. To determine the global transcriptional effects of Mi-E318K, we performed pan-genomic expression profiling of A375 melanoma and RCC4 cells infected with adenoviruses that encoded either wild-type MITF or Mi-E318K (Supplementary Fig. 6a, b, c, d). Analysis of the genes regulated differentially by wild type and Mi-E318K in

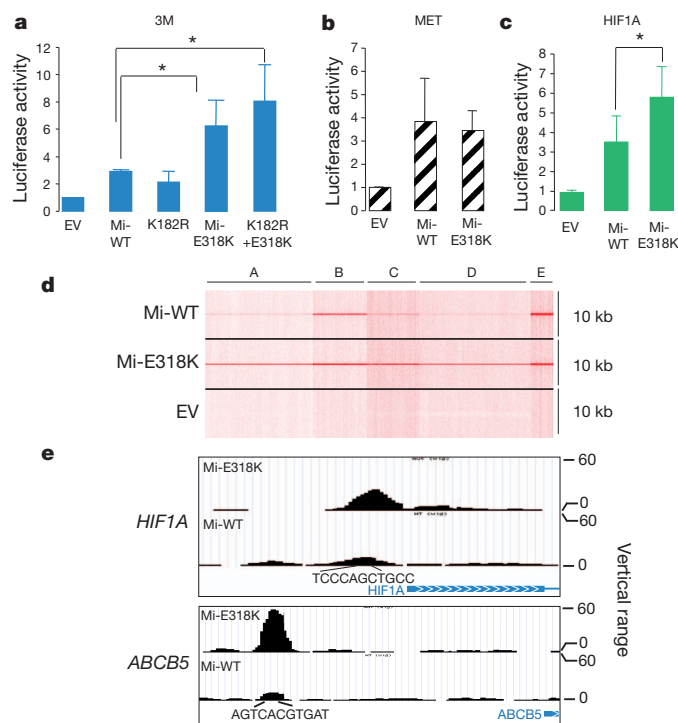


Figure 2 | Mi-E318K increases MITF binding and transcriptional regulation of a subset of its target genes. **a–c**, Reporter plasmids containing the luciferase gene under the control of different MITF target promoters were co-expressed with empty pcDNA3 (EV) or pcDNA3 encoding wild-type (Mi-WT) or mutant (K182R, Mi-E318K, K182R+E318K) MITF. **a**, HEK293 cells were transfected with a 3M-box reporter. **b**, **c**, 501mel human melanoma cells were transfected with the *c-MET* promoter reporter (**b**) or the *HIF1A* promoter reporter (**c**). The results were expressed as the fold stimulation over basal luciferase activity. Data represent the mean + s.d. of three independent experiments. Asterisk indicates a statistically significant difference (Student's *t*-test, $P \leq 0.05$). **d**, MITF-occupied loci in genomic DNA extracted from 501mel melanoma cells expressing Mi-WT or Mi-E318K. MITF tag density was compared in the region of ± 5 kb around the MITF-occupied loci. **e**, University of California Santa Cruz (UCSC) view of MITF occupancy of *HIF1A* and *ABCB5* loci. The blue bar with arrowheads below each panel indicates the transcription start site; the E-box nucleotide sequence is shown under peaks.

A375 cells failed to identify a specific Mi-E318K signature. However, in RCC4 cells, we identified a Mi-E318K signature composed of 32 genes (Supplementary Table 4). On the basis of the bioinformatic analysis using Ingenuity Pathways Analysis (IPA) software, these genes were associated with cell growth, proliferation and inflammation (Supplementary Fig. 7). Among the genes downregulated by Mi-E318K compared to wild type, *IRAK2* (ref. 11), *EGR1* (ref. 12) and *IL6* (ref. 13) have tumour suppressor functions. The upregulated genes included *CCR7*, a HIF-1 α target gene¹⁴ and activator of the NF κ B pathway¹⁵; *ABCB5*, one of the most well-documented markers for melanoma initiating cells¹⁶; *GADD45G*, a member of the GADD45 family of stress sensors, which connects NF κ B to the MAP kinase pathway¹⁷; and *TRIM63*, which was one of the top five upregulated genes and was the major gene found to be differentially expressed in patients with TFE3/TFEB translocation RCC as compared to other subtypes of RCC⁸. These experiments confirmed the different transcriptional potentials of wild type and Mi-E318K, particularly on genes involved in proliferation and inflammation.

To investigate whether SUMOylation influenced genome-wide MITF occupancy of its target sites, we performed chromatin immunoprecipitation coupled to high-throughput sequencing (ChIP-seq) experiments in 501mel human melanoma cell lines expressing wild type or Mi-E318K. Data analysis shows that 22,157 sites are occupied by Mi-E318K, whereas only 9,107 were detected for wild type, the difference being the presence of a large number of weakly occupied sites in the Mi-E318K data set. Comparative analysis of the ChIP-seq data sets using read density matrix clustering identified sites that are occupied by both wild type and Mi-E318K (Fig. 2d, sets B and E), whereas some promoter sites are uniquely or preferentially occupied by Mi-E318K (Fig. 2d, sets A and D; Supplementary Table 5). For example, a higher occupancy of the *HIF1A* promoter by Mi-E318K compared to wild type is observed (Fig. 2e). Mi-E318K also showed increased binding to the *ABCB5* locus (Fig. 2e), the expression of which is increased in RCC4-expressing Mi-E318K. Of note, Mi-E318K binds more efficiently than wild type to the *HMOX1* promoter (Supplementary Table 5), a gene involved in both kidney cancer¹⁸ and melanoma cell growth¹⁹. Although the *HIF1A* promoter was better occupied and activated (in a gene promoter assay) by Mi-E318K than by wild type, we observed equal amounts of *HIF1A* messenger RNA in Mi-E318K- or wild-type-infected A375 and RCC4 cells. *HIF1A* transcripts are subjected to post-transcriptional regulation that controls their stability²⁰; this may mask the transcriptional effect of Mi-E318K. Alternatively, the cellular context or the chromatin landscape in RCC4 or A375 cells may not permit increased transcription. Indeed, this was observed for MITF target genes such as *TYR*, *TYRP1* and *DCT*, which were not upregulated by MITF overexpression in A375 cells. Taken together, our data indicate that the naturally occurring Mi-E318K severely impaired MITF SUMOylation and showed higher global transcriptional activity, in agreement with the currently accepted model of SUMO-mediated transcriptional repression²¹. The global increase in Mi-E318K-occupied loci coupled with the existence of sites exclusively bound by the mutant protein indicate that SUMOylation regulates the repertoire of MITF target genes. SUMOylation-deficient Mi-E318K protein may therefore result in the regulation of distinct sets of genes, hence leading to gain-of-function properties.

To evaluate the tumorigenic potential of Mi-E318K, we tested its effect on migration, invasion and colony formation of stable melanoma cells (501mel) expressing wild-type MITF or Mi-E318K (Fig. 3a). We observed increased migration (Fig. 3b) and invasion (Fig. 3c) in two different Mi-E318K clones compared to wild-type clones. Similar results were obtained in VHL-deficient RCC4 cells (Fig. 3d, e) infected with adenoviruses encoding wild type or Mi-E318K. In contrast, Mi-E318K caused barely significant effects in A375 cells (used also for expression profiling) (Supplementary Fig. 8). Lastly, in a colony-forming assay with immortalized murine melanocytes, Melan-a (Fig. 3f), and RCC4 cells (Fig. 3g), the number of colonies after transfection of

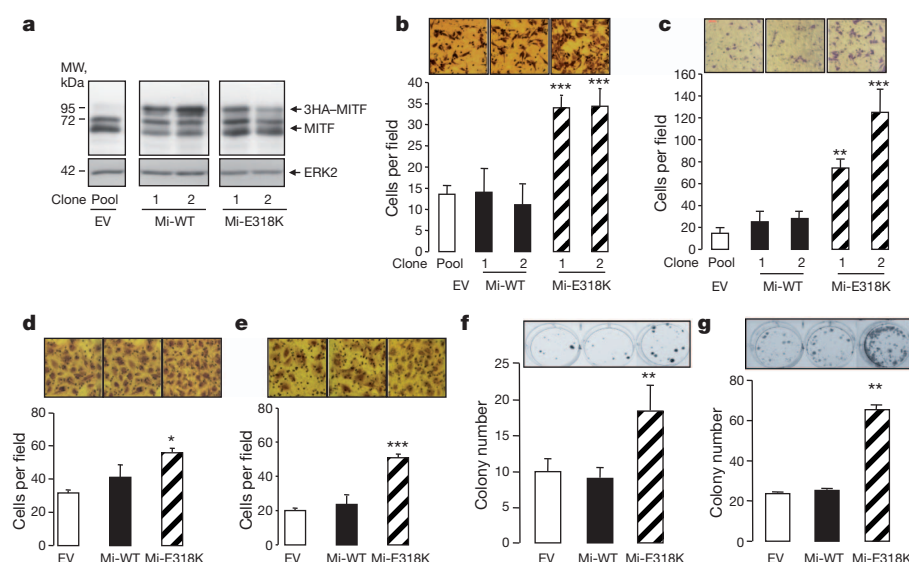


Figure 3 | Mi-E318K-enhanced migration, invasion and clonogenicity of melanoma and renal cancer cells. **a**, Western blot analysis with antibodies to MITF or ERK2 of 501mel cells stably transfected with empty vector (EV), or vector encoding Mi-WT or Mi-E318K (clones 1 and 2). MW, molecular weight. **b**, **c**, Migration (**b**) and invasion (**c**) assays with 501mel melanoma cells described earlier. **d**, **e**, Migration and invasion (**e**) assays in RCC4 cells transduced with empty adenovirus (EV) or adenoviruses encoding either Mi-WT or Mi-E318K. **b–e**, Photographs above each bar graph show a representative field of the underside of filter after cell migration or invasion; magnification, $\times 200$. **f**, **g**, Mi-E318K increased the number of colonies formed with Melan-a or RCC4 cells, respectively. Photographs of the colonies were taken at 2 weeks. **b–g**, Bars show the mean \pm s.d. of three replicate assays. **b–g**, Student's *t*-test compared Mi-WT to Mi-E318K. **P* < 0.05, ***P* < 0.01 and ****P* < 0.001.

Mi-E318K was enhanced compared to wild-type-transfected or control cells. Overall, these data indicated that Mi-E318K was more potent than wild type for promoting invasive and tumorigenic behaviours in melanoma and RCC cells. However, Mi-E318K did not significantly stimulate melanoma or RCC cell growth (data not shown). These features are reminiscent of melanoma cell populations with increased invasive and division potential but with a slow growth rate, which are considered to be melanoma-initiating cells²². Thus, Mi-E318K might favour a phenotypic switch of melanoma cells towards a tumour-initiating cell phenotype, possibly in synergy with hypoxia²³.

Indeed, both the epidermis (where melanocytes are located) and the inner renal medulla are physiologically hypoxic tissues^{24,25}. Hypoxia is capable of rapidly inducing HIF1A to initiate a cell-survival response²⁶. Of note, several genes identified by expression profiling or ChIP-seq (*HIF1A*, *CCR7*, *HMOX1*) function in the hypoxia pathway. Both ultraviolet exposure and hypoxia generate radical oxygen species (ROS). In turn, oxidative stress activates inflammatory pathways that can lead to cellular transformation, proliferation and stem cell survival, among other features²⁷. Our IPA analysis associated the Mi-E318K signature with inflammation, cell proliferation and cancer. Various environmental stresses, including hypoxia and ROS, were previously shown to induce global protein SUMOylation²⁸. In this context, Mi-E318K, which prevents MITF SUMOylation, could impair the adaptation of cells to stress and initiate tumour formation. Our study has wide implications for understanding the role of MITF and its SUMOylation in the physiology and tumorigenesis of melanocytes and kidney cells.

We have identified a rare oncogenic germline substitution, Mi-E318K, that predisposes to both melanoma and RCC. Interestingly, *MITF* loss-of-function mutations are responsible for an inherited disorder in neural crest cell development, the type 2a Waardenburg syndrome⁷. This is reminiscent of the *RET* oncogene, in which activating germline mutations predispose to medullary thyroid carcinoma, but loss-of-function germline mutations predispose to Hirschsprung's disease, a congenital absence of enteric neurons in the gastrointestinal tract²⁹. Our data indicate that mutation screening in genetically enriched patients is a powerful strategy to identify rare genetic variation that confers a moderate risk of cancer³⁰. Lastly, it highlights the relevance of hereditary tumour models for shedding light on cell-growth-related signals and identifying cancer driver genes.

METHODS SUMMARY

Participants were included with the approval of an institutional review board. Sanger sequencing of *MITF* isoforms was performed in 62 melanoma + RCC cases. Mutations numbering referred to NCBI accession NM_000248.3. Wild-type or Mi-E318K genotyping was performed in 1,659 controls and 829 cases by TaqMan; positive samples were checked by Sanger sequencing. Genome-wide

SNP data were genotyped by the Centre National de Génotypage (CNG, CEA) using Illumina technology. PCA was applied to the SNP data using EIGENSTRAT software. Immunohistochemistry was performed on melanoma and kidney fixed samples with anti-MITF antibody (Abcam, clone D5). Mutations in the two SUMO1 fixation sites, p.K182R and p.E318K, were generated with the QuickChange method in wild-type *MITF-M* and *MITF-A* isoforms and verified by sequencing. Plasmids were transfected into HEK293, 501mel, Melan-a and RCC4 cells with FuGENE 6 reagent (Roche Applied Science). Western blotting and immunofluorescence were performed with antibodies to MITF (Abcam, clone C5), HA-tag, SUMO1 or ERK2 (Santa Cruz Biotechnology). Reporter assays were performed by transient transfection in triplicate with lipofectamine (Invitrogen), and pCMVβGal was included to control transfection efficiency. Gene expression microarray assays were performed with Agilent technology by the Institut de Cancérologie Gustave Roussy (IGR) genomics platform. ChIP-seq experiments were performed by the Institut de Génétique et de Biologie Moléculaire et Cellulaire (IGBMC) microarray and sequencing platforms. Colony forming assays were performed after 14 days of culture; cells were fixed, stained with 0.4% crystal violet, and photographed. Statistical analyses for comparison of allele frequencies between cases and controls were performed with Fisher's exact test. All computations were performed with Stata software, version 11. For functional assays, statistical significance was evaluated with the Student's *t*-test. Results were considered significant when the Student's *t*-test, **P* value ≤ 0.05 , ***P* ≤ 0.01 , ****P* ≤ 0.001 .

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 1 February; accepted 2 September 2011.

Published online 19 October 2011.

1. Tucker, M. A. Melanoma epidemiology. *Hematol. Oncol. Clin. North Am.* **23**, 383–395 (2009).
2. Rini, B. I., Campbell, S. C. & Escudier, B. Renal cell carcinoma. *Lancet* **373**, 1119–1132 (2009).
3. Maubec, E. *et al.* Characteristics of the coexistence of melanoma and renal cell carcinoma. *Cancer* **116**, 5716–5724 (2010).
4. Garraway, L. A. *et al.* Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117–122 (2005).
5. Cheli, Y., Ohanna, M., Ballotti, R. & Bertolotto, C. 15-year quest in search for MITF target genes. *Pigment Cell Melanoma Res.* **23**, 27–40 (2009).
6. Linehan, W. M., Srinivasan, R. & Schmidt, L. S. The genetic basis of kidney cancer: a metabolic disease. *Nature Rev. Urol.* **7**, 277–285 (2010).
7. Hershey, C. L. & Fisher, D. E. Genomic analysis of the microphthalmia locus and identification of the MITF-J/MITF-J isoform. *Gene* **347**, 73–82 (2005).
8. Camparo, P. *et al.* Renal translocation carcinomas: clinicopathologic, immunohistochemical, and gene expression profiling analysis of 31 cases with a review of the literature. *Am. J. Surg. Pathol.* **32**, 656–670 (2008).
9. Granter, S. R., Weillbaecher, K. N., Quigley, C. & Fisher, D. E. Role for microphthalmia transcription factor in the diagnosis of metastatic malignant melanoma. *Appl. Immunohistochem. Mol. Morphol.* **10**, 47–51 (2002).
10. Murakami, H. & Arnheiter, H. Sumoylation modulates transcriptional activity of MITF in a promoter-specific manner. *Pigment Cell Res.* **18**, 265–277 (2005).
11. Mullenders, J. *et al.* Interleukin-1R-associated kinase 2 is a novel modulator of the transforming growth factor β signaling cascade. *Mol. Cancer Res.* **8**, 592–603 (2010).
12. Yu, J. *et al.* PTEN regulation by Akt–EGR1–ARF–PTEN axis. *EMBO J.* **28**, 21–33 (2009).

13. Wysocki, P. J. *et al.* Gene-modified tumor vaccine secreting a designer cytokine Hyper-Interleukin-6 is an effective therapy in mice bearing orthotopic renal cell cancer. *Cancer Gene Ther.* **17**, 465–475 (2010).
14. Li, Y., Qiu, X., Zhang, S., Zhang, Q. & Wang, E. Hypoxia induced CCR7 expression via HIF-1 α and HIF-2 α correlates with migration and invasion in lung cancer cells. *Cancer Biol. Ther.* **8**, 322–330 (2009).
15. Liu, F. Y. *et al.* NF- κ B participates in chemokine receptor 7-mediated cell survival in metastatic squamous cell carcinoma of the head and neck. *Oncol. Rep.* **25**, 383–391 (2011).
16. Schatton, T. *et al.* Identification of cells initiating human melanomas. *Nature* **451**, 345–349 (2008).
17. Yang, Z., Song, L. & Huang, C. Gadd45 proteins as critical signal transducers linking NF- κ B to MAPK cascades. *Curr. Cancer Drug Targets* **9**, 915–930 (2009).
18. Datta, D., Banerjee, P., Gasser, M., Waaga-Gasser, A. M. & Pal, S. CXCR3-B can mediate growth-inhibitory signals in human renal cancer cells by down-regulating the expression of heme oxygenase-1. *J. Biol. Chem.* **285**, 36842–36848 (2010).
19. Was, H. *et al.* Overexpression of heme oxygenase-1 in murine melanoma: increased proliferation and viability of tumor cells, decreased survival of mice. *Am. J. Pathol.* **169**, 2181–2198 (2006).
20. Wang, M. J. & Lin, S. A region within the 5'-untranslated region of hypoxia-inducible factor-1 α mRNA mediates its turnover in lung adenocarcinoma cells. *J. Biol. Chem.* **284**, 36500–36510 (2009).
21. Garcia-Dominguez, M. & Reyes, J. C. SUMO association with repressor complexes, emerging routes for transcriptional control. *Biochim. Biophys. Acta* **1789**, 451–459 (2009).
22. Hoek, K. S. & Goding, C. R. Cancer stem cells versus phenotype-switching in melanoma. *Pigment Cell Melanoma Res.* **23**, 746–759 (2010).
23. Li, Z. & Rich, J. N. Hypoxia and hypoxia inducible factors in cancer stem cell maintenance. *Curr. Top. Microbiol. Immunol.* **345**, 21–30 (2010).
24. Zou, A. P. & Cowley, A. W. Jr. Reactive oxygen species and molecular regulation of renal oxygenation. *Acta Physiol. Scand.* **179**, 233–241 (2003).
25. Bedogni, B. & Powell, M. B. Skin hypoxia: a promoting environmental factor in melanomagenesis. *Cell Cycle* **5**, 1258–1261 (2006).
26. Bellot, G. *et al.* Hypoxia-induced autophagy is mediated through hypoxia-inducible factor induction of BNIP3 and BNIP3L via their BH3 domains. *Mol. Cell. Biol.* **29**, 2570–2581 (2009).
27. Reuter, S., Gupta, S. C., Chaturvedi, M. M. & Aggarwal, B. B. Oxidative stress, inflammation, and cancer: how are they linked? *Free Radic. Biol. Med.* **49**, 1603–1616 (2010).
28. Tempé, D., Piechaczyk, M. & Bossis, G. SUMO under stress. *Biochem. Soc. Trans.* **36**, 874–878 (2008).
29. Manié, S., Santoro, M., Fusco, A. & Billaud, M. The RET receptor: function in development and dysfunction in congenital malformation. *Trends Genet.* **17**, 580–589 (2001).
30. Manolio, T. A. *et al.* Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank the patients and family members who participated in this study and the clinicians who identified these families, the French Familial Melanoma Study Group and the Inherited Predisposition to Kidney Cancer network. We acknowledge the contribution of the IGR Biobank for providing MELARISK samples and the CEPH Biobank for processing DNA samples. We thank L. Larue, J. Feunteun, A. Sarasin and E. Solary for critical reviews of the manuscript. We thank V. Lazar and S. Forget for coordination of the IGR's genomics and genetic platforms, N. Pata-Merci, V. Marty, S. Le Gras and A. Chabrier for their technical expertise, and M. Barrois for technical counselling. We also thank A. Boland for DNA extraction and quality control for genome-wide genotyping. This work was supported by grants from INSERM, Ligue Nationale Contre Le Cancer (PRE05/FD and PRE 09/FD) to F.D.; Programme Hospitalier de Recherche Clinique (PHRC 2007/AOM-07-195) to M.-F.A. and F.D.; ARC N°A09/5/5003 to B.B.-d.P.; ARC 4985 to C.B.; Institut National du Cancer (INCa)-Cancéropole Ile de France (melanoma network RS#13) to B.B.-deP.; INCa-PNCS rein to B.G., S.Ga. and S.R.; INCa grant R08009AP to C.B.; Fondation de France 2010 to R.B.; INCa and Ligue Nationale Contre Le Cancer to I.D.; Fond de maturation IGR and Fondation Gustave Roussy to B.B.-d.P.; Société Française de Dermatologie SDF2004 to R.B. and P.B.; SFD2009 to B.B.-d.P.; 2009 SGR 1337 from AGAUR, Generalitat de Catalunya, and FIS PS09/01393 from the Fondo de Investigaciones Sanitarias, Instituto de Salud Carlos III, Spain to S.P. and C.B.; and personal donations from C. and N. de Paillerets and M.-H. Wagner to B.B.-d.P. B.B.-d.P. holds an INSERM Research Fellowship for hospital-based scientists. Work at the Centre National de Génotypage (CNG) and Centre d'Etude du Polymorphisme Humain (CEPH) was supported in part by INCa.

Author Contributions C.B., F.L., M.L., F.D., R.B. and B.B.-d.P. designed the experiments and wrote the manuscript. A.Re., B.G., S.S. and G.M.L. participated in the scientific discussions. E.M., P.Va., S.D., N.P., T.M.-D., L.T., P.A.-B., N.D., F.B., A.Ro., J.-L.P., B.L., C.R., B.E., O.C., L.B., S.R., J.C., B.T., P.Gh., L.P., S.P., C.B., H.O., C.I., E.R., R.L. and P.B. collected biological samples. P.Ga. collected the control samples. F.L., M.d.L. and B.d'H. performed sequencing and genotyping of patients. H.B. supervised DNA extraction and quality control for genome-wide genotyping. D.Z. and M.L. were responsible for the genome-wide genotyping of cases and controls and genotyping of MITF variant in controls. E.C. carried out the analysis of SNP genotype data. F.D. supervised the statistical analysis of all genotyped data. K.B. and S.Gi. performed the functional analysis. A.d.I.F., V.M. and P.Vi. performed MITF immunostaining. T.S. and I.D. designed and performed the ChIP-seq experiments. P.D. performed the gene expression

profiling analysis. M.-F.A. initiated the collection of melanoma and RCC cases. S.R. initiated the collection of RCC families. M.-F.A. and F.D. initiated the MELARISK collection. H.M. and V.C. contributed to the management of the MELARISK database.

Author Information Genome data has been deposited at the European Genome-Phenome Archive (EGA; <http://www.ebi.ac.uk/ega>), which is hosted at the EBI, under accession number EGAS00000000048. Gene expression data related to this paper have been submitted to the Array Express repository at the European Bioinformatics Institute (<http://www.ebi.ac.uk/arrayexpress/>) under the accession number E-TABM-1198. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.B.-d.P. (brigitte.bressac@igr.fr).

The French Familial Melanoma Study Group

François Aubin¹, Bertrand Bachollet², Céline Becuwe³, Pascaline Berthet⁴, Yves Jean Bignon⁵, Valérie Bonadona⁶, Jean-Louis Bonafe⁷, Marie-Noëlle Bonnet-Dupeyron⁸, Frédéric Cambazard⁹, Jacqueline Chevrant-Breton¹⁰, Isabelle Coupier¹¹, Sophie Dalac¹², Liliane Demange¹³, Michel d'Incan¹⁴, Catherine Dugast¹⁵, Laurence Faivre¹⁶, Lynda Vincent-Fétita¹⁷, Marion Gauthier-Villars¹⁸, Brigitte Gilbert¹⁹, Florent Grange²⁰, Jean-Jacques Grob²¹, Philippe Humbert¹, Nicolas Janin²², Pascal Joly²³, Delphine Kerob²⁴, Christine Lasset⁶, Dominique Leroux²⁵, Julien Levang¹, Jean-Marc Limacher²⁶, Cristina Livideanu²⁷, Michel Longy²⁸, Alain Lortholary²⁹, Dominique Stoppa-Lyonnet¹⁸, Sandrine Mansard¹⁴, Ludovic Mansuy³⁰, Karine Marrou³, Christine Matéus², Christine Maugard³¹, Nicolas Meyer³², Catherine Nogues³³, Pierre Souteyrand¹⁴, Laurence Venat-Bouvét³⁴ & Hélène Zattara³⁵

¹Centre Hospitalier Universitaire St Jacques, Dermatologie et Vénéréologie, 2 Place St Jacques, 25030 Besançon, France. ²Institut de Cancérologie Gustave Roussy, Service de Dermatologie, 114 Rue Edouard Vaillant, 94805 Villejuif, France. ³Hôpital de l'Hôtel-Dieu, Service de Dermatologie, 1 place de l'Hôpital, 69288 Lyon Cedex 2, France. ⁴Service d'Oncologie Génétique, Centre François Baclesse, 3 Avenue du Général Harris, 14076 Caen Cedex 5, France. ⁵Centre Hospitalier Jean Perrin, 58 rue Montalembert, BP 392, 63011 Clermont-Ferrand Cedex 1, France. ⁶Centre Léon Bérard, Unité Clinique d'Oncologie Génétique, 28 Rue Laënnec, 69373 Lyon Cedex 8, France. ⁷Unité de Dermatologie, Centre Hospitalier Universitaire de Rangueil, 1 avenue Jean Poulhès, TSA 50032, 31059 Toulouse, France. ⁸Centre Hospitalier, Service de Génétique, 179 Boulevard Maréchal Juin, 26953 Valence Cedex 9, France. ⁹Centre Hospitalier Universitaire de St Etienne, Hôpital Nord, Service de Dermatologie, Vénéréologie, 42055 St Etienne Cedex 2, France. ¹⁰Chru Pontchaillou, Service de Dermatologie, 35 Rue André Le Guilloux, 35033 Rennes, France. ¹¹Centre Hospitalier Universitaire Hôpital Arnaud de Villeneuve, Unité d'Oncogénétique, 371 Avenue Doyen Gaston Giraud, 34295 Montpellier Cedex 5, France. ¹²Centre Hospitalier Universitaire Hôpital du Bocage, Service de Dermatologie, 2 boulevard Maréchal de Lattre de Tassigny, BP 77908, 21079 Dijon, France. ¹³Polyclinique Courlancy, Service de Radiothérapie et Oncologie Médicale, 38 rue de Courlancy, 51000 Reims, France. ¹⁴Centre Hospitalier Universitaire Estaing, Service de Dermatologie, 1 Place Lucie Aubrac, 63003 Clermont Ferrand Cedex 1, France. ¹⁵Centre Eugene Marquis, Oncologie Génétique, Rue de la Bataille de Flandres Dunkerque, CS 44229, 35042 Rennes, France. ¹⁶Centre Hospitalier Universitaire Hôpital du Bocage, Centre de Génétique, 2 Boulevard Maréchal de Lattre de Tassigny, BP77908, 21079 Dijon, France. ¹⁷Hôpital Cochin, Service de Dermatologie, Pavillon Tarnier, 89 rue d'Assas, 75006 Paris, France. ¹⁸Institut Curie, Génétique Oncologique, 26 rue d'Ulm, 75248 Paris Cedex 5, France. ¹⁹Centre Hospitalier Universitaire La Milétrie, Service de Génétique Médicale, 2 rue de la Milétrie, BP 577, 86021 Poitiers, France. ²⁰Hôpital Robert Debré, Service de Dermatologie U42, Avenue du Général Koehn, 51092 Reims, France. ²¹Hôpital Sainte-Marguerite, Service de Dermatologie, 270 Boulevard Sainte-Marguerite, 13274 Marseille Cedex 9, France. ²²Centre Hospitalier Universitaire Sart Tilman, Département de Génétique Humaine, Centre Hospitalier Universitaire de Liege, 4000 Liege, Belgium. ²³Centre Hospitalier Universitaire de Rouen, Hôpitaux de Rouen, Clinique Dermatologique, Hôpital Charles Nicolle, 1 rue de Germont, 76031 Rouen, France. ²⁴Hôpital Saint-Louis, Service de Dermatologie, 1 Avenue Claude Vellefaux, 75475 Paris Cedex 10, France. ²⁵Centre Hospitalier Universitaire de Grenoble, Hôpital Couple Enfant, Département de Génétique et Procréation, Génétique Clinique, Consultations d'Oncogénétique, BP 217-F, 38043 Grenoble Cedex 9, France. ²⁶Hôpital Pasteur, Hôpitaux civils de Colmar, Service d'Oncologie, 39 Avenue Liberté, 68024 Colmar, France. ²⁷Centre Hospitalier Universitaire Toulouse, Hôpital Purpan, Service de Dermatologie, Place du Docteur Baylac TSA 40031, 31059 Toulouse Cedex 9, France. ²⁸Institut Bergonié, Laboratoire de Génétique Moléculaire, 229 cours de l'Argonne, 33076 Bordeaux, France. ²⁹Centre Catherine de Sienne, Oncologie Médicale, 2 rue Eric Tabarly, BP 20215, 44202 Nantes Cedex 2, France. ³⁰Centre Alexis Vautrin, Consultation Oncogénétique, UF Nancy 9901, 6 avenue de Bourgogne, 54511 Vandoeuvre Les Nancy, France. ³¹Centre Hospitalier Universitaire de Strasbourg, Consultations d'Oncogénétique, Service d'Hématologie et d'Oncologie, 1 place de l'Hôpital, 67091 Strasbourg, France. ³²Centre Hospitalier Universitaire de Toulouse, Hôpital Larrey, Service de Dermatologie, Vénéréologie, 24 Chemin de Pouvoirville, TSA 30030, 31059 Toulouse Cedex 9, France. ³³Centre René Huguenin, Service d'Oncogénétique, 35 rue Dailly, 92210 St Cloud, France. ³⁴Oncogénétique, Centre Hospitalier Universitaire de Limoges, Hôpital Universitaire de Dupuytren, 2 Avenue Martin Luther King, 87042 Limoges, France. ³⁵Département de Génétique Médicale, Hôpital de la Timone, Unité de Génétique Clinique, 264, rue St Pierre, 13005 Marseille, France.

METHODS

Study participants. Patients with melanoma and RCC. Sixty-two patients that had developed both melanoma and RCC were included in the MELARISK collection at the Institut de Cancérologie Gustave Roussy (IGR) and other French hospitals, 71% were males. None of these patients carried a germline *CDKN2A*, *CDK4* or *VHL* pathogenic mutation.

Patients with melanoma only. The patients with melanoma were enrolled through a nationwide network of French dermatology and oncogenetic clinics that constituted the Familial Melanoma Project and the MELARISK collection. The present study investigated a total of 371 index cases from independent pedigrees with a family history of melanoma (at least two melanoma cases) and/or pancreatic cancer, and 232 sporadic cases diagnosed with multiple primary malignant melanomas. A subset of the families ($N = 34$) included a melanoma patient and relatives with pancreatic cancer. These were part of the *CDKN2A* mutation testing procedure; carriers of the *CDKN2A* mutation have high risk of developing pancreatic cancer³¹. All cases had confirmed diagnoses of malignant melanoma through medical records, review of pathological material, and/or pathological reports. For all these patients, mutation screening for *CDKN2A* (exon 1β, 1α, 2 and 3) and *CDK4* (exon 2) had been undertaken for molecular diagnosis purposes. None of these patients carried a known *CDKN2A* or *CDK4* pathogenic mutation.

Patients with RCC only. Overall, 164 patients included in this study had developed RCC only. Those patients had been recruited through French urology and oncogenetic clinics, within the French National Cancer Institute's 'Inherited Predisposition to Kidney Cancer' network. They were considered 'genetically enriched' based on the following criteria: familial aggregation, young age of onset and rare histological subtypes. Out of the total of 164 patients, 79 had sporadic ccRCC, 54 had papillary renal cell carcinoma (PRCC; with 6 type I PRCC, 19 type II PRCC, 1 mixed type PRCC and 28 unknown histological subtypes of PRCC), 1 had a mixed RCC phenotypes (that is, papillary and ccRCC), 5 had juvenile RCC and 25 had RCC of unknown histological subtype. None of the 79 patients with ccRCC carried a *VHL* pathogenic germline mutation; none of the 19 patients with PRCCII carried a *FH* germline mutation; and none of the 6 patients with PRCCI carried a *MET* germline mutation.

Control subjects. Controls were supplemented by a population-based sample of 1,659 French subjects that had participated in the Supplementation in Vitamins and Mineral Antioxidants (SU.VI.MAX) study³². We confirmed that these subjects had not developed any cancer at the time of the present study.

The study was approved by an institutional review board (CCPRB no. 01-09-05, Paris Necker) for the MELARISK collection and by the Ethical Committee of Le Kremlin-Bicêtre University Hospital, for the kidney cancer collection. It was conducted according to the Declaration of Helsinki Principles. All participating subjects signed informed consent and provided blood samples. After informed consent had been obtained from the subjects, DNA was extracted from peripheral blood lymphocytes with the QIAamp DNA Blood mini kit (QIAGEN), according to manufacturer guidelines.

MITF gene sequencing. Sanger sequencing of *MITF* coding sequence, the intron-exon boundaries and the 8 alternative promoters was performed in DNA extracted from the blood of 62 patients affected with melanoma + RCC (primers available upon request) on a 3730 DNA Analyser (Applied Biosystems; ABI). Nucleotide number refers to the wild-type cDNA sequence of *MITF* (NM_000248.3) as reported in GenBank.

Genotyping of Mi-E318K. Genotyping was carried out with Taqman according to the manufacturer recommendations. Primers and probes were supplied in the Assay-by-Design by Applied Biosystems (ABI). PCR reactions were performed with 10 ng genomic DNA and 0.2 μmol l⁻¹ TaqMan MGB probes. Probe 5'-VIC-ATCAAGCAAGAACC CG-3' was designed to match the wild-type allele, and probe 5'-6-FAM-CAAGCAAAAACCCG-3' was designed to match the mutant allele (underlined bold indicates the nucleotide at the mutation site). PCR thermocycling was performed on ABI thermocyclers, as follows: 95 °C for 15 min; 30 cycles of 95 °C for 15 s and 60 °C for 1 min. Assays were carried out in 96-well plates that included a negative control (with no DNA) and positive controls (DNA from subjects with Mi-E318K). Plates were read on a 7900HT Fast Real-Time PCR System (ABI) with Sequence Detection Software (ABI). Carriers of Mi-E318K were confirmed by Sanger sequencing with the above protocol and primers for *MITF* exon 9.

PCA of genome-wide SNP data. To verify that the association of Mi-E318K with melanoma and/or RCC was not influenced by population stratification, we carried out PCA of SNP data across the genome in cases and controls that had been genome-wide genotyped and satisfied stringent quality control criteria. A total of 1,628 controls and 569 cases affected with melanoma, RCC or both cancers were genotyped by the Centre National de Génotypage (CNG) using Illumina HumanHap300 Beadchip version 2 duo array for controls and Illumina Humanncv370k and Human660W-Quad arrays for cases. Samples were excluded

for any of the following reasons: (1) a call rate of less than 97% of the total number of SNPs on the chip (222 samples); (2) sex as ascertained by genotyping not matching reported sex (10 samples); (3) heterozygosity on autosomes departing from the estimated expected value (7 samples); (4) relatedness with another sample (4 samples). This resulted in exclusion of 243 samples. PCA was then applied to 1,954 subjects that passed quality control using EIGENSTRAT software³³. To identify individuals of non-European ethnicity, SNPs were thinned to reduce linkage disequilibrium and combined with the HapMap data of wide-ranging ethnicity. The first two principal components (PCs) clearly separated the HapMap data into distinct clusters according to ethnicity and identified 29 study samples of non-European ancestry who were excluded. The remaining 1,925 European ethnicity samples (1,389 controls, 536 cases) were analysed similarly without the Hapmap data. Plotting the first two principal components showed appropriate clustering of cases and controls (Supplementary Fig. 1).

MITF immunohistochemistry. Briefly, 4-μm tissue sections were cut from paraffin-embedded blocks, deparaffinized and rehydrated. RCC (patient ID numbers 10276, 24976, 11473, 21309, 21939, 25220, 26534; all in Supplementary Table 2) immunohistochemistry was performed with a three-step, avidin-biotin-peroxidase method. Staining was performed in an automated Dako Autostainer. Heat-mediated antigen retrieval was performed in 0.1 mol l⁻¹ pH 6.0 citrate, and samples were heated in a water bath for 30 min in a Dakolink processor (Dako). Monoclonal mouse anti-human MITF (Clone D5, Dako, 1/100) was incubated with the samples for 30 min at room temperature (22 °C). SSM immunohistochemistry (patient ID numbers 19525, 24976, 10254, 15168, 15012, 21000, 22112, 27708 all in Supplementary Table 2 except number 27708 in Supplementary Fig. 1) was performed with a three-step avidin-biotin-alkaline-phosphatase method. Staining was performed in a Roche Diagnostics Benchmark XT automated stainer. Heat-mediated antigen retrieval was performed in the automat in cell conditioning solution 1 (Roche Diagnostics) for 1 h (Dako). Monoclonal mouse anti-Human MITF (Clone D5, Roche Diagnostics, pre-diluted) was incubated with samples for 16 min at room temperature. All slides were then counterstained with haematoxylin. Appropriate negative and positive controls were prepared in parallel.

Functional assays. Plasmids. The pcDNA3-Mi construct, which carried the M-MITF isoform and the 3M vector have been described previously³⁴. The A-MITF isoform was a gift from H. Arnheiter (NIH). The *MITF* mutations, p.K182R and/or Mi-E318K, were generated with the QuickChange method (Stratagene) with the following sense primers and their reverse complements: Mi-K182R 5'-CTTCCCAACATAAGAAGGAGCTCACAGC-3'; Mi-E318K 5'-GGATCATCAAGCAAAAACAGTTCTTGAG-3'. Mutations were confirmed by DNA sequencing. His-SUMO1 and HA-SUMO2 were a gift from A. Dejean and have been described elsewhere³⁵.

Cell cultures, transfections, and immunoblots. Human 501mel and A375 melanoma cells, human RCC4 cells and HEK293 cells were grown in DMEM supplemented with 7% FBS. Melan-a cells were cultured in RPMI 1640, 7% FCS, 200 nM TPA and 200 pM cholera toxin. All the cells were maintained at 37 °C in a humidified atmosphere containing 5% CO₂. HEK293 cells were cultured in 6-well dishes (10⁴ cells per well) and transfected with the above-mentioned plasmids (2 μg of total DNA per well) and FuGENE 6 (Roche Applied Science). Forty-eight hours later, the cells were washed with PBS and lysed at 95 °C in 1× loading buffer (41.6 mM Tris, pH 6.8, 1.5% SDS, 6.7% glycerol). Proteins were resolved by electrophoresis in 10% SDS-polyacrylamide gels and transferred to PVDF membranes. Proteins were detected with ECL (Amersham) and antibodies to MITF (Abcam), HA-tag (Abcam), SUMO1 (Santa Cruz Biotechnology) or ERK2 (Santa Cruz Biotechnology).

Reporter assays. Human 501mel melanoma cells and HEK293 cells were seeded in 24-well dishes (25 × 10³ cells per well). Subsequently, cells were transiently transfected with 0.3 μg reporter plasmid (3M, pHIF1A, pMET), 0.05 μg of MITF-encoding plasmids or empty, control pcDNA3 (EV), 2 μl of lipofectamine reagent (Invitrogen) and 0.05 μg of pCMVβGal for controlling the variability in transfection efficiency. Cells were lysed 48 h later and assayed for luciferase and β-galactosidase activities. Transfections were repeated at least three times.

Immunofluorescence. HEK293 cells were seeded on glass coverslips (100 × 10³ cells) in 6-well dishes, and transfected with 3 μg of the different MITF mutants and 10 μl lipofectamine. Forty-eight hours later, cells were fixed for 10 min with 4% paraformaldehyde, permeabilized for 2 min with 0.1% Triton X-100/1% bovine serum albumin (BSA), and treated for 2 min with NH₄Cl 50 mM. Then, samples were washed three times in PBS and stained for 1 h with monoclonal anti-MITF antibody (Abcam, clone D5) in 1% BSA/PBS. Next, samples were washed three times in PBS for 5 min each and then stained secondarily for 1 h with Alexa-594-conjugated goat-anti-mouse antibody (Molecular probes) in 1% BSA. Cells were counterstained with 4,6-diamidino-2-phenylindole (DAPI), mounted with Fluoromount-G (Southern Biotech), and examined with a Zeiss Axiophot microscope equipped with epifluorescent illumination.

Colony forming assay. Immortalized Melan-a mouse melanocytes, A375 human melanoma cells and RCC4 human cells (8×10^4 cells per well) were transfected with wild type or Mi-E318K ($3 \mu\text{g}$ total DNA per well and 10% pBABE-puro) with the FuGENE 6 Transfection reagent (Roche Applied Science). Puromycin ($1 \mu\text{g ml}^{-1}$) was added to media at 48 h post-transfection. Fourteen days later, the cells were fixed and stained with 0.4% crystal violet, and the plates were photographed.

Establishing stable 501mel cells that expressed tagged MITF. 501mel cells were transfected with FuGENE 6 reagent, a vector encoding puromycin resistance, and a pCMV vector that was either an empty, control vector or a vector encoding 3 \times HA-tagged MITF (wild type or Mi-E318K). Transfected cells were selected with puromycin; the expression of MITF was verified by western blot analysis with anti-MITF (Abcam, C5) or anti-HA (Roche, 12CA5) antibodies.

Migration and invasion assays. Migration (on uncoated filters) and invasion (on coated filters with matrigel) were investigated in a Boyden chamber system that comprised 24-well plates and 8- μm pore filter inserts (BD Bioscience). Stable 501mel melanoma cells (2×10^5 cells), A375 melanoma cells (1.5×10^5 cells) and RCC4 renal cancer cells (5×10^4 cells) were infected with control vector or adenovirus that encoded wild type or Mi-E318K at a multiplicity of infection (MOI) of 20 for 24 h, were then resuspended in serum-free DMEM and seeded on the upper chamber inserts. DMEM with 7% FCS was placed into the lower chamber. Twenty-four hours (501mel and RCC4) or 6 h (A375 cells) later, cells adherent to the underside of the filters were fixed with 4% PFA and stained with 0.4% crystal violet. Adherent cells were counted in five random fields at $\times 200$ magnification. Results represented the average of duplicate samples from three independent experiments.

Gene expression arrays. Three replicates were performed for both RCC4 and A375 cells. mRNA was isolated with Trizol (Invitrogen) from A375 melanoma cells and from RCC4 cells infected with either control, wild type or Mi-E318K, according to standard procedures. Briefly, probes were synthesized from 500 ng of total RNA in two steps, according to the manufacturer's instructions. The two samples to be compared were labelled separately with different fluorescent dyes, cyanine-3 (Cy3) and cyanine-5 (Cy5). For each sample, $1 \mu\text{g}$ of purified cRNA labelled in cy5 was mixed with the same amount of cRNA labelled in cy3. Label incorporation was checked on a NanoDrop spectrophotometer. Hybridizations were performed with a dye-swap strategy on whole-human-genome dual colour $8 \times 60\text{K}$ oligonucleotide microarrays (design 028004; Agilent Technologies). Feature extraction software provided by Agilent (version 10.7.3.1) was used to quantify the intensity of fluorescent images and to apply a linear/lowess normalization to correct for artefacts caused by nonlinear rates of dye incorporation and inconsistent relative fluorescence intensities between some green and red dyes. All data were imported into Resolver software (Rosetta Biosoftware) for database management, quality control, computational re-combination of dye-swaps and statistical analysis. Mi-E318K specific signature was generated using the following parameters: intensity > 50 ; fold change > 1.5 ; P value < 0.05 . Functional analysis was carried out with the Ingenuity Pathway Analysis (Ingenuity System, <http://www.ingenuity.com>).

ChIP-seq. ChIP-seq was performed as previously described³⁶. Briefly, chromatin was prepared from native 501mel cells or cells that stably expressed HA-tagged wild type or Mi-E318K. The isolated chromatin was HA-immunoprecipitated and sequenced with an Illumina GAIIX sequencer. The raw data were analysed with the Illumina Eland pipeline V1.6 programme. Peak detection was performed with MACS software (<http://liulab.dfci.harvard.edu/MACS/>), and the peaks were annotated with GPAT software (http://bips.u-strasbg.fr/GPAT/Gpat_home.html). Peak annotations were performed with a window that included $\pm 20\text{ kb}$ from the coordinates at the beginning and end of RefSeq transcripts. The total number of reads for the wild-type data set was 1.7 fold higher than for the Mi-E318K data set. To facilitate quantitative comparisons, an appropriate number of reads was randomly removed from the wild-type data set to match the number present in the Mi-E318K data set. Subsequently, a quantitative comparison of the ChIP-seq data sets was performed with seqMINER³⁷. Clustering was performed by counting the number of tags in a 36-bp sliding window for each ChIP-seq data set based on the coordinates of wild-type or Mi-E318K binding sites or on a reference sequence (RefSeq TSS). Only peaks with ≥ 10 reads were used for comparison. A matrix of binding sites and densities was generated and subjected to K-means clustering with the seqMINER programme.

Statistical analysis. Fisher's exact test was used to compare Mi-E318K allele frequencies between cases and controls and among different groups of cases. The odds ratios and confidence intervals associated with Mi-E318K carrier status were estimated by exact logistic regression. Logistic regression was also used to test for the effect of Mi-E318K on melanoma and/or RCC with and without adjusting for the first ten principal components estimated from genome-wide SNP data in patients and controls who were genome-wide genotyped and passed quality control. All computations were performed with Stata software, version 11 (StataCorp LP). For functional assays, results were considered significant when the Student's t -test, $*P$ value was ≤ 0.05 , $**P \leq 0.01$, $***P \leq 0.001$.

31. Goldstein, A. M. *et al.* High-risk melanoma susceptibility genes and pancreatic cancer, neural system tumors, and uveal melanoma across GenoMEL. *Cancer Res.* **66**, 9818–9828 (2006).
32. Hercberg, S. *et al.* The SU.VI.MAX Study: a randomized, placebo-controlled trial of the health effects of antioxidant vitamins and minerals. *Arch. Intern. Med.* **164**, 2335–2342 (2004).
33. Patterson, N., Price, A. L. & Reich, D. Population structure and eigenanalysis. *PLoS Genet.* **2**, e190 (2006).
34. Bertolotto, C., Bille, K., Ortonne, J. P. & Ballotti, R. Regulation of tyrosinase gene expression by cAMP in B16 melanoma cells involves two CATGTG motifs surrounding the TATA box: implication of the microphthalmia gene product. *J. Cell Biol.* **134**, 747–755 (1996).
35. Bischof, O. *et al.* The E3 SUMO ligase PIASy is a regulator of cellular senescence and apoptosis. *Mol. Cell* **22**, 783–794 (2006).
36. Martianov, I. *et al.* Cell-specific occupancy of an extended repertoire of CREM and CREB binding loci in male germ cells. *BMC Genomics* **11**, 530 (2010).
37. Ye, T. *et al.* seqMINER: an integrated ChIP-seq data interpretation platform. *Nucleic Acids Res.* **39**, e35 (2011).

A novel recurrent mutation in *MITF* predisposes to familial and sporadic melanoma

Satoru Yokoyama^{1*}, Susan L. Woods^{2*}, Glen M. Boyle^{2*}, Lauren G. Aoude^{2*}, Stuart MacGregor^{2*}, Victoria Zismann^{3*}, Michael Gartside², Anne E. Cust⁴, Rizwan Haq¹, Mark Harland⁵, John C. Taylor⁵, David L. Duffy², Kelly Holohan², Ken Dutton-Regester², Jane M. Palmer², Vanessa Bonazzi², Mitchell S. Stark², Judith Symmons², Matthew H. Law², Christopher Schmidt², Cathy Lanagan², Linda O'Connor², Elizabeth A. Holland⁶, Helen Schmid⁶, Judith A. Maskiell⁷, Jodie Jetann⁸, Megan Ferguson⁸, Mark A. Jenkins⁷, Richard F. Kefford⁶, Graham G. Giles⁹, Bruce K. Armstrong⁴, Joanne F. Aitken⁸, John L. Hopper⁷, David C. Whiteman², Paul D. Pharoah¹⁰, Douglas F. Easton¹⁰, Alison M. Dunning¹⁰, Julia A. Newton-Bishop⁵, Grant W. Montgomery², Nicholas G. Martin², Graham J. Mann⁶, D. Timothy Bishop⁵, Hensin Tsao^{11*}, Jeffrey M. Trent^{3*}, David E. Fisher^{1*}, Nicholas K. Hayward^{2*} & Kevin M. Brown^{3,12*}

So far, two genes associated with familial melanoma have been identified, accounting for a minority of genetic risk in families. Mutations in *CDKN2A* account for approximately 40% of familial cases¹, and predisposing mutations in *CDK4* have been reported in a very small number of melanoma kindreds². Here we report the whole-genome sequencing of probands from several melanoma families, which we performed in order to identify other genes associated with familial melanoma. We identify one individual carrying a novel germline variant (coding DNA sequence c.G1075A; protein sequence p.E318K; rs149617956) in the melanoma-lineage-specific oncogene microphthalmia-associated transcription factor (*MITF*). Although the variant co-segregated with melanoma in some but not all cases in the family, linkage analysis of 31 families subsequently identified to carry the variant generated a log of odds (lod) score of 2.7 under a dominant model, indicating E318K as a possible intermediate risk variant. Consistent with this, the E318K variant was significantly associated with melanoma in a large Australian case-control sample. Likewise, it was similarly associated in an independent case-control sample from the United Kingdom. In the Australian sample, the variant allele was significantly over-represented in cases with a family history of melanoma, multiple primary melanomas, or both. The variant allele was also associated with increased naevus count and non-blue eye colour. Functional analysis of E318K showed that *MITF* encoded by the variant allele had impaired sumoylation and differentially regulated several *MITF* targets. These data indicate that *MITF* is a melanoma-predisposition gene and highlight the utility of whole-genome sequencing to identify novel rare variants associated with disease susceptibility.

Cutaneous malignant melanoma is predominantly a disease of fair-skinned individuals. Aetiology is complex, with environmental (mainly ultraviolet radiation exposure) and genetic factors affecting disease risk. Phenotypic risk factors, which are largely heritable, include pigmentation (fair skin, blue or green eyes, blonde or red hair), sun sensitivity, an inability to tan³⁻⁶, high number of melanocytic naevi^{7,8}, or the presence of clinically atypical naevi⁷. Candidate-gene studies and genome-wide association studies (GWAS) for melanoma and these melanoma-associated phenotypes have identified several

variants associated with melanoma risk in the general population⁹⁻¹³. Family studies, on the other hand, have identified only two high-penetrance melanoma genes, *CDKN2A* (ref. 1) and *CDK4* (ref. 2), accounting for a minority of genetic risk in melanoma families.

As part of a larger sequencing effort to identify novel melanoma risk genes, we sequenced the genome of an affected individual from an eight-case melanoma family negative for alterations in *CDKN2A* or *CDK4* (Fig. 1, FAM1) using a nanoarray-based short-read sequencing-by-ligation strategy¹⁴. From among the 410 novel variants predicted to affect protein structure, we prioritized for follow-up a single nucleotide polymorphism (SNP) resulting in a glutamic acid to lysine substitution

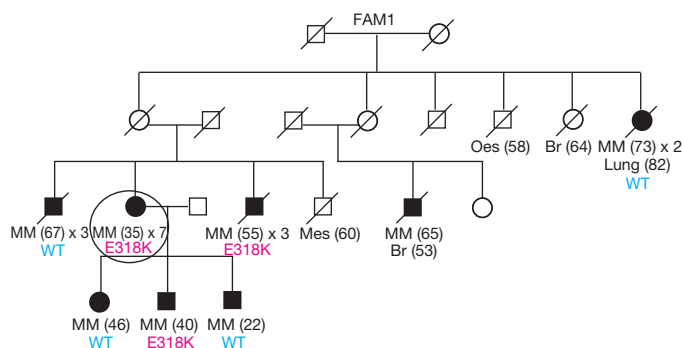


Figure 1 | Co-segregation analysis of the *MITF* E318K variant in the family in which it was identified. The pedigree shows individuals that have had melanoma (shaded circles or boxes), with the age of first melanoma diagnosis indicated in brackets and the number of melanomas that have occurred in the individual so far (for example, × 2 indicates two primary melanomas). If the number of melanomas is not stated, the individual has had a single melanoma. A diagonal line through the symbol indicates that the person is deceased. The genotype for the *MITF* E318K variant for individuals with an available DNA sample for testing is annotated 'E318K' if a carrier or wild type 'WT'. Other cancer types are also indicated with the age of first diagnosis indicated in brackets if known. Br, breast; Mes, mesothelioma; MM, melanoma; Oes, oesophagus. The individual circled in Family 1 (FAM1) is the melanoma case in which the *MITF* E318K variant was discovered through whole-genome sequencing. See Supplementary Fig. 3 for pedigrees of all other families identified as carrying E318K.

¹Department of Dermatology, Cutaneous Biology Research Center, Massachusetts General Hospital, Harvard Medical School, Massachusetts 02114, USA. ²Queensland Institute of Medical Research, Brisbane, Queensland 4029, Australia. ³Translational Genomics Research Institute, Phoenix, Arizona 85004, USA. ⁴Cancer Epidemiology and Services Research, Sydney School of Public Health, Sydney Medical School, The University of Sydney, New South Wales 2006, Australia. ⁵Section of Epidemiology and Biostatistics, Leeds Institute of Molecular Medicine, Cancer Research UK Clinical Centre at Leeds, St James's University Hospital, Leeds LS9 7TF, UK. ⁶Westmead Institute of Cancer Research, University of Sydney at Westmead Millennium Institute and Melanoma Institute Australia, Westmead, New South Wales 2145 Australia. ⁷Centre for Molecular, Environmental, Genetic and Analytic Epidemiology, School of Population Health, University of Melbourne, Victoria 3052 Australia. ⁸Viertel Centre for Research in Cancer Control, The Cancer Council Queensland, Brisbane, Queensland 4004 Australia. ⁹Cancer Epidemiology Centre, The Cancer Council Victoria, Carlton, Victoria 3053 Australia. ¹⁰University of Cambridge, Cambridge CB1 8RN, England, United Kingdom. ¹¹Wellman Center for Photomedicine, Massachusetts General Hospital, Harvard Medical School, Massachusetts 02114, USA. ¹²Division of Cancer Epidemiology and Genetics, National Cancer Institute, National Institutes of Health, Gaithersburg, Maryland 20892, USA.

*These authors contributed equally to this work.

Table 1 | *MITF* E318K association with melanoma

Population	Group	No. of individuals with variant	No. of individuals without variant	Variant carrier frequency	OR	95% CI	One-sided exact <i>P</i>	Two-sided exact <i>P</i>
Australia	Cases	34	2,025	0.0165	2.33	1.21–4.70	0.0045	0.0083
	Controls	14	1,939	0.0072				
UK	Cases	34	1,895	0.0176	2.09	1.14–3.94	0.0074	0.0115
	Controls	18	2,097	0.0085				
Australia + UK	Cases	68	3,920	0.0171	2.19	1.41–3.45	0.0001	0.0003
	Controls	32	4,036	0.0079				

It should be noted that the reported allele frequencies for *MITF* E318K in the population-based samples are without removing individuals with *CDKN2A* or *CDK4* mutations, as screening for these genes was not routinely performed.

in *MITF* (E318K, codon numbering based on the melanocyte-specific *MITF*-M isoform; c.G1075A, NCBI accession NM_000248.3; p.E318K, NCBI accession NP_000239.1; rs149617956). Although linkage¹⁵ and GWAS studies^{9,10} have not provided evidence implicating *MITF* in either predisposition to melanoma or the melanoma-associated phenotypes of pigmentation and naevogenesis^{11,12,16–19}, *MITF* is known to regulate a broad repertoire of genes whose functions in melanocytes range from development, differentiation, survival, cell-cycle regulation and pigment production. *MITF* is somatically amplified^{20,21} or mutated²² in a subset of melanomas, and strongly over-expressed in others²⁰, making it an attractive candidate despite the lack of prior evidence for involvement in germline risk.

We evaluated whether *MITF* E318K is a high-penetrance melanoma susceptibility variant in Family 1 by genotyping the remaining affected individuals available for study. The *MITF* variant allele was found in 3/7 melanoma cases assessed in this family (Fig. 1), consistent with it being a medium-penetrance melanoma risk variant. To assess further this possibility, we genotyped two large Australian melanoma case-control samples for *MITF* E318K. The variant was found in 14/1,953 controls (carrier frequency = 0.0072) and thus represents a rare population variant (Table 1). We observed a significantly higher frequency (34/2,059) in cases (carrier frequency = 0.0165) than controls (Fisher exact *P* = 0.008, odds ratio (OR) 2.33, 95% confidence interval (CI) 1.21–4.70), indicating that the variant correlates with increased melanoma risk in the general population. The effect size for E318K is larger than those reported for variants from melanoma GWAS^{9,10} and similar to that observed for red-hair-colour-associated variants of the melanocortin 1 receptor (*MC1R*) gene (OR for most populations ~2.4)²³. Among cases, the *MITF* E318K variant was enriched in those with multiple primary melanomas (OR 4.22, 95% CI 1.52–10.91), a family history of melanoma (OR 2.95, 95% CI 1.23–6.92), or both (OR 8.37, 95% CI 2.58–23.80), but not in cases with earlier age of onset (comparing diagnosis before age 40 versus after 40 years) (Table 2).

We replicated these findings in two independent population-based case-control samples from the United Kingdom. In the combined UK sample, the variant allele frequency was also significantly higher in cases (carrier frequency = 0.0176) than controls (carrier frequency = 0.0085, *P* = 0.012, OR 2.09, 95% CI 1.14–3.94, Table 1). The association with melanoma in the pooled UK and Australian data was highly significant (combined *P* = 0.0003, OR 2.19, 95% CI 1.41–3.45). In the UK cases there were also trends towards family history, earlier age of onset, and the occurrence of more than one primary melanoma in variant carriers (Table 2).

To extend assessment of the *MITF* variant in melanoma-prone families, we screened for E318K in 182 UK families with at least two melanoma cases and 88 Australian families with at least three cases, all of which are negative for mutations in *CDKN2A* or *CDK4*. Six families (2.2%) were found to carry the variant. In the UK, E318K was enriched in the more melanoma-dense families; 4/54 (7.4%) families with at least three melanoma cases versus 1/128 (0.8%) families with two melanoma cases (Fisher's exact *P* = 0.013). We subsequently evaluated whether *MITF* E318K co-segregated with melanoma in these as well as additional multiple-case families identified from the case-control sample. In total, we identified 31 unrelated cases carrying *MITF* E318K from Australia and the UK with at least one first- or second-degree relative diagnosed with melanoma (listed in Supplementary Table 1; Supplementary Fig. 3), 22 of which had DNA available from additional affected family members for genotyping. In 9/31 families (five three-case and four two-case families) the variant was found in all affected individuals (Supplementary Fig. 3a; non-segregating families shown in Supplementary Fig. 3b), whereas in 12 additional families, the variant co-segregated with melanoma in the available cases, but DNA from all cases was not available for screening (Supplementary Fig. 3c). To test more formally for linkage of melanoma with *MITF* E318K in these families, we calculated a lod score of 2.7 under a dominant model, again consistent with an incompletely penetrant medium risk variant.

To assess whether the *MITF* variant is related to known melanoma-associated risk phenotypes of pigmentation and naevus count, we tested for association both in cases and controls from the Australian and British populations. The *MITF* variant allele is significantly associated with increased naevus count (combined *P* = 0.002, OR 2.54, 95% CI 1.42–4.55; Supplementary Table 2) and non-blue eye colour (combined *P* = 0.018, OR 2.01, 95% CI 1.11–3.81; Supplementary Table 3). It was not associated with skin colour, hair colour, or freckling (Supplementary Table 4). Reassessing the case-control analysis accounting for naevus count and eye colour gave a slightly reduced effect size for association of the variant with melanoma (OR 1.82, 95% CI 0.85–3.92), suggesting that the risk of melanoma attributable to *MITF* E318K may be mediated at least in part via one or both of these phenotypes, but that there is a substantial residual risk conferred by the variant through an as yet undetermined mechanism.

We next sought to evaluate whether the E318K mutation alters *MITF* function. The E318K variant is located within one of two IKXE consensus sites on *MITF* previously shown to be post-translationally modified by the addition of the small ubiquitin-like-modifier SUMO²⁴.

Table 2 | Association of *MITF* E318K with melanoma-associated variables

Case set	Australia				UK			
	Carrier frequency	Carrier counts (carrier/wild type)	OR vs Aus controls	95% CI	Carrier frequency	Carrier counts (carrier/wild type)	OR vs UK controls	95% CI
All cases	0.0165	34/2,025	2.33	1.21–4.70	0.0176	34/1,895	2.09	1.14–3.94
Age of onset <40 years	0.0139	14/996	1.95	0.86–4.42	0.0174	5/283	2.10	0.60–5.91
Family history of melanoma	0.0209	12/563	2.95	1.23–6.92	0.0273	3/107	3.36	0.62–11.77
Multiple primary melanomas	0.0296	8/262	4.22	1.52–10.91	0.0225	2/87	2.74	0.30–11.74
Multiple melanomas and family history	0.0571	6/99	8.37	2.58–23.80	0.0000	0/10	-	-

Aus, Australia.

Mutation of the residue to which SUMO is covalently attached in this motif (K316R) has previously been shown to abrogate MITF sumoylation and significantly increase MITF transcriptional activity *in vitro*^{24,25}. We thus hypothesized that E318K would similarly alter sumoylation and transcriptional activity of MITF. To test this we constructed a cDNA encoding His-tagged MITF carrying the E318K mutation. We evaluated the effects of E318K on sumoylation in comparison to the wild type and previously characterized synthetic mutations of the two known MITF sumoylation sites (K316R and K182R) by co-transfecting with haemagglutinin (HA)-tagged SUMO1 in COS-7 cells (Fig. 2a). Wild-type MITF shows two SUMO1-modified forms, whereas MITF mutants K182R or K316R each show only one modified form (Fig. 2a). Similar to the synthetic K316R and K182R mutants, E318K abrogates sumoylation, resulting in complete loss of the doubly sumoylated form of MITF and reducing the mono-sumoylated form. When the second site is mutated (K182R) simultaneously with E318K, MITF sumoylation is completely abolished. Immunoprecipitation of endogenously expressed MITF E318K from melanoma cells homozygous for E318K (NAE) when transfected with SUMO similarly revealed only bands corresponding to mono- and non-sumoylated isoforms of MITF on western blot (Fig. 2b).

We then looked for differences between mutant and wild-type MITF transcriptional activity using a reporter construct containing the MITF-responsive *TRPM1* promoter²⁵. At two concentrations tested, the E318K mutant exhibited 1.34–1.40 fold induction of the *TRPM1* luciferase reporter relative to wild-type MITF (Fig. 2c). This fold induction is similar to that observed previously on multiple MITF target promoters using single or double artificial sumoylation-site MITF mutants^{24,25} and suggests that the E318K variant found in melanoma changes the transcriptional potency of MITF. To study this in greater detail, we determined the effect of the E318K mutation on global MITF target gene transcription. We developed a tetracycline-inducible system for expression of wild-type MITF or the E318K variant in melanoma cell lines with constitutively low or undetectable levels of endogenous MITF (HT144 and C32, respectively²⁶, Fig. 2d). At the phenotypic level, induction of wild-type or E318K MITF led to increased proliferation compared to uninduced controls for each of the cell lines, although there was no significant difference in growth rate between the cells expressing the different isoforms (data not shown). We examined whole-genome expression profiles in these cells following induction of either wild-type or E318K MITF for 48 h. Of the 37 genes commonly regulated by wild-type and E318K MITF in both cell lines (Supplementary Fig. 1a, b; see Methods for analysis details), 28 (76%) had previously been identified as MITF target genes (Supplementary Table 5)^{27,28}, and 17 showed ≥ 1.25 -fold differences in expression between the wild-type and E318K isoforms (Supplementary Fig. 1b). We also identified two gene products that were uniquely differentially regulated compared to uninduced cells by the induction of wild-type MITF but not MITF E318K in both parental cell lines, and 16 gene products after induction of MITF E318K but not wild-type MITF (Supplementary Table 6). Of these, 61% (11/18) have previously been reported as MITF targets (Supplementary Table 6)^{27,28}. Collectively, these data indicate that the MITF E318K mutant exhibits differential transcriptional activity against some, although not all, target genes. In agreement with the reporter assays (Fig. 2c), we identified transcriptional differences in gene products known to be involved in pigmentation (*DCT*, *MLANA*), in which the differences were more marked with expression of the E318K variant in comparison to wild-type MITF. These were validated by quantitative polymerase chain reaction with reverse transcription (qRT-PCR) in the cell lines used for microarray analysis (Fig. 2e), as well as in an additional melanoma cell line constitutively expressing wild-type or E318K mutant MITF (Fig. 2f and Supplementary Fig. 2). In keeping with the increase in expression of these pigmentation genes, we detected a 22% increase in melanin content in HT-144 melanoma cells 72 h after induction of MITF E318K compared to wild-type MITF

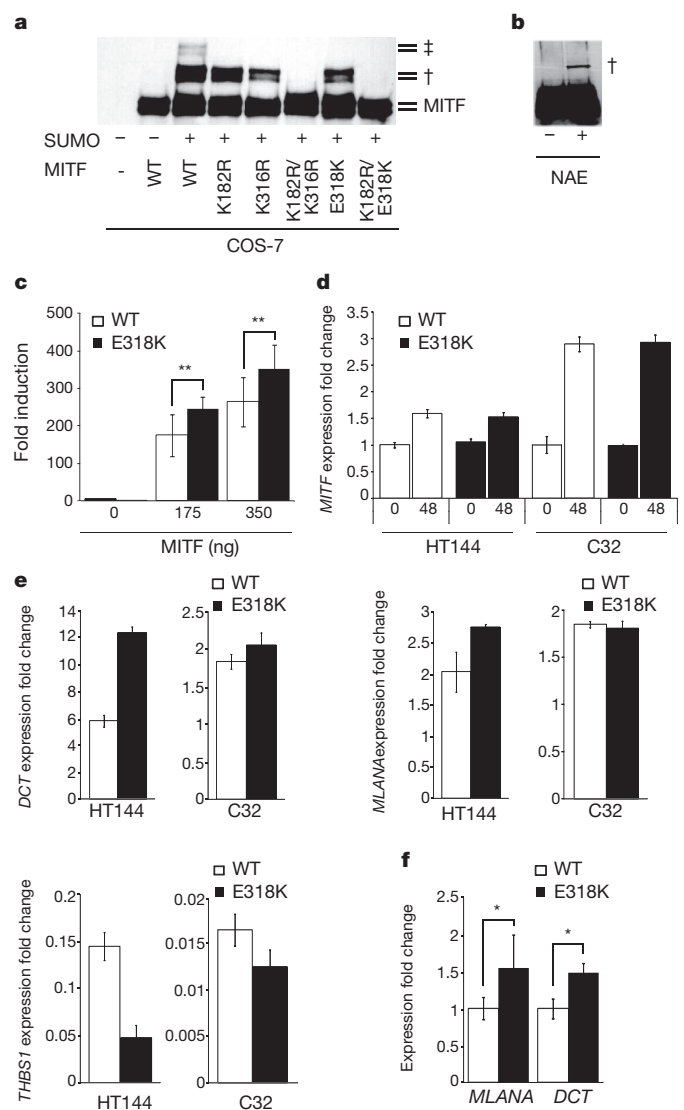


Figure 2 | E318K prevents MITF sumoylation and results in differential expression of MITF target genes. **a**, His-tagged wild-type MITF or the indicated single or double point mutants were co-transfected with HA-SUMO1 in COS-7 cells or **b**, HA-SUMO1 was transfected alone into homozygous mutant E318K MITF melanoma cells (NAE). Single- and double-sumoylated forms of MITF are indicated by a dagger and double dagger, respectively. The doublet bands are caused by MAPK-mediated phosphorylation at serine 73 (ref. 30). **c**, UACC62 human melanoma cells were transfected with *TRPM1*-promoter constructs with indicated amounts of expression vector encoding wild-type or mutant forms of MITF. Fold induction is shown as the ratio to the average of no MITF transfection (0 ng). Data are mean \pm s.d. of at least four independent experiments. **d**, Expression of MITF in two melanoma cell lines (HT144 and C32) engineered to inducibly express wild-type (WT) or mutant (E318K) MITF after treatment with tetracycline for 48 h (48), as determined by qRT-PCR. Performed in triplicate, error bars depict s.d. **e**, Expression of MITF target genes *DCT* (top left), *MLANA* (top right) and *THBS1* (bottom left) determined by qRT-PCR in melanoma cell lines 48 h after induction of wild-type or E318K MITF. Gene expression is normalized to *GAPDH* and shown as fold change compared to uninduced cells. Performed in triplicate, error bars denote s.d. **f**, qRT-PCR analysis of total RNA isolated from UACC62 human melanoma cells, which were transfected with expression vector encoding wild-type or mutant forms of MITF. The expression level of each target gene was normalized to MITF mRNA. Fold induction is shown as the ratio to each mRNA expression with wild-type MITF. Data are mean \pm s.d. of at least three independent experiments. * $P < 0.05$, ** $P < 0.01$.

(data not shown). This is also consistent with our observation that carriers are more likely to have darker (that is, non-blue) eye colour (Supplementary Table 3) but, notably, these data contrast with other

previously reported 'fair-skin-associated' melanoma risk variants, such as those in *MC1R* or *TYR*. It is uncertain whether the enhanced expression of pigment genes may contribute to melanomagenesis, perhaps by increasing oxidative stress and an increase in oxidative DNA damage²⁹, or alternatively may simply reflect increased *MITF* activity, which (separately) promotes tumorigenesis, as *MITF* is a previously recognized amplified melanoma oncogene²⁰.

We adopted the approach of whole-genome sequencing of patients from melanoma families and identified a novel germline mutation of *MITF*. This mutation was found to be present in numerous melanoma families, as well as the general population, in which its association with melanoma has an effect size similar to red-hair-causing variants of *MC1R*²³. The melanoma susceptibility genes discovered through GWAS so far account for only a minority of inherited disease risk. A proportion of this 'missing heritability' may be due to rare sequence variants, which are poorly detected by GWAS using SNP arrays. The new *MITF* variant reported here shows reasonably strong linkage to melanoma (lod score 2.7) but crucially not a high enough signal to be clearly visible in previous genome-wide linkage scans. We also provide *in vitro* data supporting a functional mechanism by which this mutation may mediate melanoma risk, specifically abrogation of *MITF* sumoylation and differential transcription of select *MITF* target genes. Although the individual changes in transcription induced by the mutant E318K *MITF* in comparison to wild-type *MITF* are modest, the orchestrated change in the levels of multiple *MITF* target genes is likely to be biologically important, especially over the lifetime of a person. This study offers a rare glimpse of a complex functionality whereby a risk-conferring SNP affects the post-translational processing of a crucial lineage-specific survival and differentiation gene. This study demonstrates the utility of performing whole-genome and exome resequencing in appropriate affected individuals to identify such novel rare disease-specific variants and functionally characterize variants associated with complex disease not otherwise detectable via GWAS or linkage approaches.

METHODS SUMMARY

The collection of the Australian melanoma families used for the study, as well as the Queensland and AMFS case-control sets are described elsewhere and in Methods. Likewise the UK studies from Leeds and Cambridge as well as the panel of melanoma cell lines. Whole-genome sequencing, assembly and variant calling were performed by Complete Genomics, as described previously¹⁴. Genotyping of *MITF* E318K was performed using the Sequenom MassArray system (Australian studies) or a custom TaqMan assay (UK studies), with DNA from the affected family member in which E318K was identified included multiple times as a positive control. Statistical analyses are described in detail in Methods. Co-segregation analyses were performed in melanoma families via Sanger sequencing using the primers: forward, 5'-CAGGCTCGAGTCATGGA-3'; reverse, 5'-TGGGGACACTATAGGCTTGG-3'. *MITF* sumoylation and *TRPM1* reporter assays were performed as previously described²⁵.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 17 March; accepted 13 October 2011.

Published online 13 November 2011.

- Goldstein, A. M. *et al.* Features associated with germline CDKN2A mutations: a GenoMEL study of melanoma-prone families from three continents. *J. Med. Genet.* **44**, 99–106 (2007).
- Zuo, L. *et al.* Germline mutations in the p16INK4a binding domain of CDK4 in familial melanoma. *Nature Genet.* **12**, 97–99 (1996).
- Naldi, L. *et al.* Cutaneous malignant melanoma in women. Phenotypic characteristics, sun exposure, and hormonal factors: a case-control study from Italy. *Ann. Epidemiol.* **15**, 545–550 (2005).
- Titus-Ernstoff, L. *et al.* Pigmentary characteristics and moles in relation to melanoma risk. *Int. J. Cancer* **116**, 144–149 (2005).
- Holly, E. A., Aston, D. A., Cress, R. D., Ahn, D. K. & Kristiansen, J. J. Cutaneous melanoma in women. I. Exposure to sunlight, ability to tan, and other risk factors related to ultraviolet light. *Am. J. Epidemiol.* **141**, 923–933 (1995).
- Holly, E. A., Aston, D. A., Cress, R. D., Ahn, D. K. & Kristiansen, J. J. Cutaneous melanoma in women. II. Phenotypic characteristics and other host-related factors. *Am. J. Epidemiol.* **141**, 934–942 (1995).

- Bataille, V. *et al.* Risk of cutaneous melanoma in relation to the numbers, types and sites of naevi: a case-control study. *Br. J. Cancer* **73**, 1605–1611 (1996).
- Chang, Y. M. *et al.* A pooled analysis of melanocytic nevus phenotype and the risk of cutaneous melanoma at different latitudes. *Int. J. Cancer* **124**, 420–428 (2009).
- Bishop, D. T. *et al.* Genome-wide association study identifies three loci associated with melanoma risk. *Nature Genet.* **41**, 920–925 (2009).
- Brown, K. M. *et al.* Common sequence variants on 20q11.22 confer melanoma susceptibility. *Nature Genet.* **40**, 838–840 (2008).
- Duffy, D. L. *et al.* *IRF4* variants have age-specific effects on nevus count and predispose to melanoma. *Am. J. Hum. Genet.* **87**, 6–16 (2010).
- Falchi, M. *et al.* Genome-wide association study identifies variants at 9p21 and 22q13 associated with development of cutaneous nevi. *Nature Genet.* **41**, 915–919 (2009).
- Gudbjartsson, D. F. *et al.* *ASIP* and *TYR* pigmentation variants associate with cutaneous melanoma and basal cell carcinoma. *Nature Genet.* **40**, 886–891 (2008).
- Drmanac, R. *et al.* Human genome sequencing using unchained base reads on self-assembling DNA nanoarrays. *Science* **327**, 78–81 (2010).
- Gillanders, E. *et al.* Localization of a novel melanoma susceptibility locus to 1p22. *Am. J. Hum. Genet.* **73**, 301–313 (2003).
- Han, J. *et al.* A genome-wide association study identifies novel alleles associated with hair color and skin pigmentation. *PLoS Genet.* **4**, e1000074 (2008).
- Nan, H. *et al.* Genome-wide association study of tanning phenotype in a population of European ancestry. *J. Invest. Dermatol.* **129**, 2250–2257 (2009).
- Sulem, P. *et al.* Genetic determinants of hair, eye and skin pigmentation in Europeans. *Nature Genet.* **39**, 1443–1452 (2007).
- Zhu, G. *et al.* A genome-wide scan for naevus count: linkage to CDKN2A and to other chromosome regions. *Eur. J. Hum. Genet.* **15**, 94–102 (2007).
- Garraway, L. A. *et al.* Integrative genomic analyses identify *MITF* as a lineage survival oncogene amplified in malignant melanoma. *Nature* **436**, 117–122 (2005).
- Stark, M. & Hayward, N. Genome-wide loss of heterozygosity and copy number analysis in melanoma using high-density single-nucleotide polymorphism arrays. *Cancer Res.* **67**, 2632–2642 (2007).
- Cronin, J. C. *et al.* Frequent mutations in the *MITF* pathway in melanoma. *Pigment Cell Melanoma Res.* **22**, 435–444 (2009).
- Williams, P. F., Olsen, C. M., Hayward, N. K. & Whiteman, D. C. Melanocortin 1 receptor and risk of cutaneous melanoma: a meta-analysis and estimates of population burden. *Int. J. Cancer* **129**, 1730–1740 (2010).
- Murakami, H. & Arnheiter, H. Sumoylation modulates transcriptional activity of *MITF* in a promoter-specific manner. *Pigment Cell Res.* **18**, 265–277 (2005).
- Miller, A. J., Levy, C., Davis, I. J., Razin, E. & Fisher, D. E. Sumoylation of *MITF* and its related family members TFE3 and TFE4. *J. Biol. Chem.* **280**, 146–155 (2005).
- Boyle, G. M. *et al.* Melanoma cell invasiveness is regulated by miR-211 suppression of the *BRN2* transcription factor. *Pigment Cell Melanoma Res.* **24**, 525–537 (2011).
- Strub, T. *et al.* Essential role of microphthalmia transcription factor for DNA replication, mitosis and genomic stability in melanoma. *Oncogene* **30**, 2319–2332 (2011).
- Hoek, K. S. *et al.* Novel *MITF* targets identified using a two-step DNA microarray strategy. *Pigment Cell Melanoma Res.* **21**, 665–676 (2008).
- Smit, N. P. *et al.* Increased melanogenesis is a risk factor for oxidative DNA damage—study on cultured melanocytes and atypical nevus cells. *Photochem. Photobiol.* **84**, 550–555 (2008).
- Hemesath, T. J., Price, E. R., Takemoto, C., Badalian, T. & Fisher, D. E. MAP kinase links the transcription factor Microphthalmia to c-Kit signalling in melanocytes. *Nature* **391**, 298–301 (1998).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements This work was supported by team science awards by the Melanoma Research Alliance (J.M.T., N.K.H., H.T. and D.E.F.), the American Cancer Society (K.M.B., RSG-08-200-01), the National Institutes of Health (NIH; D.E.F., AR043369-14; N.K.H., CA88363; H.T., K24CA149202 and P50CA93683), Doris Duke Medical Foundation (D.E.F.), Dr Miriam and Sheldon G. Adelson Medical Research Foundation (D.E.F.), US-Israel Binational Science Foundation (D.E.F.), and the Division of Cancer Epidemiology and Genetics of the National Cancer Institute (K.M.B.). N.K.H., D.L.D., S.M. and G.W.M. are supported by National Health and Medical Research Council of Australia (NHMRC) research fellowships. M.H.L. is supported by Cancer Australia grant 1011143. The collection of samples in the Leeds-based case-control study (the Melanoma Cohort Study) was funded by Cancer Research UK (Project Grant C8216/A6129 and Programme awards C588/A4994 and C588/A10589) and by the NIH (R01 CA83115). Recruitment was facilitated by the UK National Cancer Research Network. We thank S. Leake, S. Haynes, S. Waseem for Leeds case-control data collection; and H. Snowden and C. Taylor from the Leeds Cancer Research UK Cancer Centre Genomics Facility for the genotyping of the UK samples. AMFS was supported by the NHMRC (project grants 566946, 107359, 211172 and program grant number 402761 to G.J.M. and R.F.K.); the Cancer Council New South Wales (project grant 77/00, 06/10), the Cancer Council Victoria and the Cancer Council Queensland (project grant 371); and the NIH (via R01 grant CA-83115-01A2 to the international Melanoma Genetics Consortium—GenoMEL). The University of Cambridge SEARCH study was supported by Cancer Research UK Programme awards (C490/A11021 and C8197/A10123). A.E.C. is the recipient of an NHMRC public health postdoctoral fellowship (520018) and a Cancer Institute NSW Early Career Development Fellowship (10/ECF/2-06). B.K.A. is supported by a University of Sydney Medical Foundation Program Grant and J.L.H. is an Australia Fellow of the NHMRC. We gratefully acknowledge all of the participants, the work and dedication of the research coordinators, interviewers, examiners and data management staff, including J. Arbuckle, S. Columbus, M. Lang,

H. Rodais, C. Ellis (Centre for MEGA Epidemiology); E. A. Holland, C. Agha-Hamilton, C. El Hayek, L. Morgan, J. Roland, E. Tyler, J. Barton, C. Watts and L. Porter (Westmead Institute of Cancer Research); M. Hillcoat, K. Holland, P. Saunders, J. Roberts and S. Tait (Viertel Centre for Research in Cancer Control); A. Kurien, C. Patterson, C. Thoo, S. de Zwaan, A. Sklavos, S. Manoharan, J. Cahill and S. Brennand (skin examiners).

Author Contributions K.M.B., N.K.H. and J.M.T. designed, analysed and managed sequencing and genotyping aspects of the study. K.M.B., S.M., S.L.W. and N.K.H. wrote the paper. D.E.F. designed functional studies and contributed to manuscript preparation. J.M.T. designed genetic studies. H.T. designed functional studies, sequenced *MITF* in melanoma cell lines and contributed to manuscript preparation. Analysis of whole-genome sequence data was performed by K.H., M.S.S., N.K.H. and K.M.B. Genotyping of *MITF* E318K in all Australian samples was performed by V.Z., with statistical analyses performed by S.M. and M.H.L. D.C.W., D.L.D., G.W.M., N.K.H. and N.G.M. oversaw collection of the Queensland samples and contributed to statistical analyses, data interpretation and manuscript preparation. G.J.M., E.A.H., H.S., J.A.M., J.J., M.F., M.A.J., R.F.K., G.G.G., B.K.A., J.F.A. and J.L.H. oversaw collection and contributed to phenotypic analyses of the AMFS study, with statistical analysis performed by A.E.C. D.T.B., J.A.N.-B., P.D.P., D.F.E. and A.M.D. designed and managed aspects of genotyping

in UK case-control studies, with genotyping performed by M.H. and statistical analyses performed by J.C.T. J.M.P. assisted in analysis of Australian melanoma pedigrees. Sequencing of *MITF* in and analysis of Australian melanoma pedigrees and cell lines were performed by L.G.A., S.L.W., M.G., K.D.-R., V.B., M.S.S., J.S., C.S., C.L. and L.O'C. S.Y. and R.H. performed the sumoylation and *MITF*-transactivation studies and contributed to manuscript preparation. G.M.B. generated inducible-*MITF* melanoma cell lines, performed phenotypic and global transcript analyses and contributed to manuscript preparation.

Author Information Expression microarray data are available through the NCBI GEO website under accession GSE31269. Data for the full genome sequenced from FAM1 has been deposited in NCBI dbGAP under accession phs000419.v1.p1. Reprints and permissions information is available at www.nature.com/reprints. This paper is distributed under the terms of the Creative Commons Attribution-Non-Commercial-Share Alike licence, and is freely available to all readers at www.nature.com/nature. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to K.M.B. (brownkm2@mail.nih.gov).

METHODS

Melanoma families. Ascertainment and clinical evaluation of Australian and UK pedigrees with familial melanoma is described elsewhere^{31–35}. In an effort to maximize the chance of identifying a novel melanoma susceptibility gene, we restricted inclusion in our study to pedigrees that were mutation negative for both *CDKN2A* and *CDK4*, by single-strand conformation polymorphism (SSCP) and/or direct sequencing.

Melanoma case-control studies. Australia. Cases were drawn from (1) a population-based sample from Queensland, unselected for age at onset (Queensland study of Melanoma: Environment and Genetic Associations; Q-MEGA³⁶) and (2) a population-based case-control family study of melanoma diagnosed before age 40 years, ascertained in Brisbane, Melbourne and Sydney (AMFS)³⁷. Two sets of Australian controls were used. First, a sample of unrelated individuals was ascertained; they were mainly parents of adolescent twins (80% of the sample), together with a smaller number of twins and their siblings, recruited through schools to participate in the Brisbane Adolescent Twin Study³⁸. Additionally, a set of controls from AMFS were available for genotyping³⁷. The AMFS and twin controls did not have a personal history of melanoma. Approval for these as well as the melanoma family studies was obtained from the Human Research Ethics Committees of Queensland Institute of Medical Research (QIMR), University of Sydney, University of Melbourne and cancer registries of New South Wales, Victoria and Queensland. Informed consent was obtained from all participants. Most samples had: questionnaire-based naevus count based on a four point scale: 'none', 'a few', 'moderate' and 'very many' (Queensland); or 'none', 'few', 'some' and 'many' for AMFS; data on pigmentation variables, including: hair colour (fair, light brown, red, dark brown, black), eye colour (blue/grey, green hazel, brown/black), skin colour (light, medium, dark) and number of freckles (none, few, some/moderate, many).

United Kingdom. The Leeds-based case-control study recruited population-based incident melanoma cases diagnosed between September 2000 and December 2006 from a geographically defined area of Yorkshire and the Northern region of the UK (63% response rate)^{9,12,39}. Cases were identified by clinicians and pathology registers and via the Northern and Yorkshire Cancer Registry and Information Service to ensure overall ascertainment. For all but 18 months of the study period, recruitment was restricted to patients with Breslow thickness of at least 0.75 mm. Controls were ascertained by contacting general practitioners to identify eligible individuals. These controls were frequency-matched with cases for age and sex from general practitioners who had also had cases as a part of their patient register. Overall there was a 55% response rate for controls. The first 960 of the cases recruited and all controls were examined by trained interviewers who performed a standardized examination of the skin, recording naevi by anatomical site and size. For subsequent cases, self-reported information on naevi was obtained. In total we utilized 1,549 cases and 495 controls from this study. We also used 380 cases and 373 controls recruited by the University of Cambridge who were genotyped in the replication series. The cases and controls were recruited as a part of the SEARCH study^{40,41}, an ongoing population-based study in Eastern England. Cases were ascertained through the Eastern Cancer Registry and Information Centre, and were aged between 18 and 70 years at diagnosis. Controls were drawn from SEARCH and EPIC-Norfolk. The control set was supplemented with 1,245 additional controls from the Wellcome Trust Blood Services control set. Details of these studies have been previously published^{40–42}. Approvals for the Leeds-based family studies, the Leeds case-control study and the Leeds Melanoma Cohort study were obtained from the National Research Ethics Service, Northern and Yorkshire Research Ethics Committee. Approvals for Cambridge case-control studies were obtained from the Cambridgeshire 4 Research Ethics Committee and the Norwich Local Research Ethics Committee. Informed consent was obtained from all participants.

Melanoma cell lines. The melanoma tumour cell lines used in this study were derived from primary cutaneous melanomas or melanoma metastases, as described previously⁴³. DNA was extracted using QIAGEN QIamp Blood Maxi kits (Qiagen) according to the manufacturer's instructions.

DNA Sequencing (whole genome). One affected member from Family 1 was sequenced by Complete Genomics using a nanoarray-based short-read sequencing-by-ligation strategy. Sequencing, genome assembly (to genome build hg18/NCBI36.1), and variant calling were performed as described previously¹⁴.

Sequenom MassARRAY genotyping of E318K. *MITF* E318K was genotyped in Australian samples as part of a multiplex assay designed using the Sequenom MassARRAY Assay Design software (version 3.0). The SNP was typed using Sequenom iPLEX chemistry on a MALDI-TOF Compact Mass Spectrometer (Sequenom). The 2.5 µl PCR reactions were performed in standard 384-well plates using 12.5 ng genomic DNA, 0.8 units of Taq polymerase (HotStarTaq, Qiagen), 500 µmol of each dNTP, 1.625 mM of MgCl₂ and 100 nmol of each PCR primer (Bioneer). Standard PCR thermal cycling conditions and post-PCR extension

reactions were carried out as described previously¹⁰. The iPLEX reaction products were desalted by diluting samples with 15 µl of water and adding 3 µl of resin. The products were spotted on a SpectroChip (Sequenom), and data were processed and analysed by MassARRAY TYPER 3.4 software (Sequenom). DNA from the Family 1 sample in which E318K was identified was included multiple times as a positive control. As an additional quality control measure, we also confirmed the genotype of multiple individuals from the Australian case-control study, particularly those with a family history of melanoma, via Sanger sequencing, identifying no genotyping errors.

TaqMan genotyping of E318K. UK samples were genotyped for E318K using a custom TaqMan assay. DNA from the Australian family member in which E318K was identified was included as a positive control. Leeds melanoma pedigrees also were re-screened as a quality control measure using high-resolution melting (HRM) primers designed to assay E318K, with the results from HRM and TaqMan matching exactly.

Statistical analyses. Parametric linkage analysis on the multiple case melanoma families was conducted assuming a dominant model of inheritance, with penetrances specified as 5%, 95%, 95% for the three genotype classes. The allele frequency at the disease and marker locus was specified to be 0.004. Computations were done using the standard (lod score) parametric linkage routine implemented in GENEHUNTER-MODSCORE⁴⁴. For melanoma case-control status, number of cases carrying the variant was compared to number of controls carrying the variant using Fisher's exact test. *P* values given are two-sided except where noted otherwise. Sub-phenotypes were tested similarly, with family history, age at onset and multiple primary melanoma carrier rates compared with that in controls. Number of blue-eyed individuals carrying the variant was compared to number of non-blue eyed individuals carrying the variant using Fisher's exact test. Meta-analysis for disease/sub-phenotype status and eye colour was done by combining the raw data. Association with mole count (high versus low where high was Australian Q-MEGA 'moderate/many', Australian AMFS 'some/many', UK '>25 moles' and low was Australian Q-MEGA 'none/few', Australian AMFS 'none/few', UK '<25 moles') was tested using Fisher's exact test for each sample individually, due to differences in the exact definition of low/high. Meta-analysis results for mole count were generated by combining the individual sample ORs (where the natural log of the odds ratio was weighted for each sample by the inverse variance of the natural log of the odds ratio). The case-control analysis corrected for naevus count and eye colour was computed by fitting a logistic regression with naevus count and eye colour categories as covariates.

Sanger sequencing for E318K. Australian pedigrees for which the proband was identified as carrying the E318K variant via Sequenom, as well as the panel of melanoma cell lines, were screened for E318K via Sanger sequencing using the following primers: forward, 5'-CAGGCTCGAGTCATGGA-3'; reverse, 5'-TGGGGACACTATAGGCTTGG-3'.

MITF sumoylation assay. His-MITF and/or HA-SUMO1 expression vectors, previously described²⁵, were transfected into COS-7 or NAE cells. The cells were harvested after three days and 10 µg total protein was applied to each lane. Western blots were performed with anti-MITF monoclonal antibody (C5).

TRPM1 reporter assay. *TRPM1* reporter plasmid (100 ng), *Renilla* luciferase vector (1 ng) (as internal control), and *MITF* expression vector (0, 175 or 350 ng) were transfected into UACC62 human melanoma cells as previously described²⁵. After 2 days, luciferase activity was determined. Transfection efficiency was normalized to *Renilla* luciferase activity. Results are reported as the average of at least four independent experiments.

Construction of inducible MITF-expressing cell lines. Stable inducible expression of MITF was achieved using the ViraPower T-REx Lentiviral Expression System (Invitrogen). Briefly, cell pools were transduced with lentiviral particles containing pLENTI6/TR, and selected with 3 µg ml⁻¹ blasticidin. Wild-type MITF was cloned into pLENTI4/TO/V5-DEST using Gateway recombination. Site-directed mutagenesis of the putative sumoylation site was carried out using a QuickChange II XL site-directed mutagenesis kit (Stratagene/Agilent Technologies), and the following primer pair: E318K change: E318K forward, 5'-GGTGAATCGGATCATCAAGCAAAAACCCGTTCTTG-3' and E318K reverse, 5'-CAAGAACGGGTTTTGCTTGATGATCCCGATTAC-3'. The negative control construct was identical except that it encodes the β-galactosidase gene (Invitrogen). Lentivirus particles were packaged in 293FT cells, before being titred using MM96L cells. Target melanoma cell lines were transduced with *MITF* and mutant particles using a multiplicity of infection of less than 1, and selected with 500 µg ml⁻¹ zeocin and 1 µg ml⁻¹ blasticidin for 3 weeks. Cells were maintained on 100 µg ml⁻¹ zeocin and 0.5 µg ml⁻¹ blasticidin for all experiments.

Induction of MITF. Cells were treated the day after seeding with 1 µg ml⁻¹ tetracycline to induce *MITF* expression, and harvested 24 or 48 h afterwards. RNA was extracted using Qiagen RNeasy mini kit as per manufacturer's instructions. RNA concentration was determined using the NanoDrop 2000 (Thermo

Scientific). For mRNA analyses 0.5 µg of total RNA was reverse transcribed using Superscript III according to the manufacturer's instructions (Invitrogen) and *MITF*, *DCT*, *MLANA*, *TBSP1* and *GAPDH* expression analysed using SYBR Green PCR Master Mix (Applied Biosystems). Primer sequences were *MITF* forward, 5'-CCAAGTACCACATACAGCAAGC-3' and *MITF* reverse, 5'-TCA TCCATCTGCATACAGGACG-3', *DCT* forward 5'-CGACTCTGATTAGTCG GAACCTCA-3' and *DCT* reverse, 5'-GGTGGTTGTAGTCATCCAAGC-3', *MLANA* forward, 5'-GAGAAAACTGTGAACCTGTGGT-3' and *MLANA* reverse, 5'-GACTGTCTGCAGAGAGTTTCTCAT-3', *THBS1* forward, 5'-CGGTCCAGACACGGACCTGC-3' and *THBS1* reverse 5'-GGCTTTGG TCTCCCGCGCTT-3', *GAPDH* forward, 5'-GGCTCTCCAGAACATCATCC CTGC-3' and *GAPDH* reverse, 5'-GGGTGTCGCTGTGAAGTCAGAG-3'. qRT-PCR was performed using a 7900HT Fast Real Time PCR System (Applied Biosystems).

Generation of constitutive MITF-expressing melanoma cell lines. pcDNA4.1-HisMax-hMITF (wild type, E318K or K182R/K316R) and pEGFP-C1 (for selection with G418) were transfected into the UACC62 melanoma cell line. After 5 days of selection with G418, total RNA was collected and *MITF* target gene transcript levels assessed by quantitative RT-PCR using the Pfaffl method⁴⁵.

Expression profiling and analysis. Biotinylated cRNA was prepared with the Illumina TotalPrep RNA Amplification Kit (Ambion). Labelled cRNA was hybridized to HumanHT-12 v4 BeadChip Arrays (Illumina), and then washed and scanned according to standard Illumina protocols. Data were extracted in GenomeStudio (Illumina) using default analysis settings and no normalization method. Resulting data were imported into GeneSpring GX v11.5 (Agilent Technologies). Expression values were normalized using quantile normalization with default settings.

Entities at least 1.5 fold up- or downregulated following induction of either wild-type or E318K MITF were determined (C32 wild type, 491; C32 E318K, 705; HT144 wild type, 273; HT144 E318K, 450). Targets similarly up- or downregulated only by E318K and not wild-type MITF in both of the melanoma cell lines were identified using Venn diagram ($n = 16$ entities). Also targets similarly up- or downregulated only by wild-type and not E318K MITF in both of the melanoma cell lines were identified using Venn diagram ($n = 2$ entities). Entities similarly up- or downregulated by induction of wild-type and E318K MITF were identified (C32 wild type and E318K, 244; HT144 wild type and E318K, 150), and those targets consistently differentially expressed in both of the cell lines identified by Venn diagram ($n = 43$ entities representing 37 gene products).

NCBI GEO (<http://www.ncbi.nlm.nih.gov/geo/>) accession numbers are as follows: GEO Series reference, GSE31269; GSM774914 HT144 wild-type uninduced; GSM774915 HT144 wild-type 48 h induced; GSM774916 HT144 E318K uninduced; GSM774917 HT144 E318K 48 h induced; GSM774918 C32 wild-type

uninduced; GSM774919 C32 wild-type 48 h induced; GSM774920 C32 E318K uninduced; GSM774921 C32 E318K 48 h induced.

Melanin quantification assay. Melanin levels were determined⁴⁶ after 72 h induction of wild-type or E318K MITF, and normalized to amounts in uninduced control cells. Values are from duplicate readings from triplicate independent experiments.

31. Kefford, R. F., Salmon, J., Shaw, H. M., Donald, J. A. & McCarthy, W. H. Hereditary melanoma in Australia. Variable association with dysplastic nevi and absence of genetic linkage to chromosome 1p. *Cancer Genet. Cytogenet.* **51**, 45–55 (1991).
32. Nancarrow, D. J. *et al.* Linkage mapping of melanoma (MLM) using 172 microsatellite markers. *Genomics* **14**, 939–947 (1992).
33. Harland, M. *et al.* Germline mutations of the *CDKN2* gene in UK melanoma families. *Hum. Mol. Genet.* **6**, 2061–2067 (1997).
34. Harland, M. *et al.* Mutation screening of the *CDKN2A* promoter in melanoma families. *Genes Chromosom. Cancer* **28**, 45–57 (2000).
35. Harland, M., Mistry, S., Bishop, D. T. & Bishop, J. A. A deep intronic mutation in *CDKN2A* is associated with disease in a subset of melanoma pedigrees. *Hum. Mol. Genet.* **10**, 2679–2686 (2001).
36. Baxter, A. J. *et al.* The Queensland Study of Melanoma: environmental and genetic associations (Q-MEGA); study design, baseline characteristics, and repeatability of phenotype and sun exposure measures. *Twin Res. Hum. Genet.* **11**, 183–196 (2008).
37. Cust, A. E. *et al.* Population-based, case-control-family design to investigate genetic and environmental influences on melanoma risk: Australian Melanoma Family Study. *Am. J. Epidemiol.* **170**, 1541–1554 (2009).
38. McGregor, B. *et al.* Genetic and environmental contributions to size, color, shape, and other characteristics of melanocytic naevi in a sample of adolescent twins. *Genet. Epidemiol.* **16**, 40–53 (1999).
39. Newton-Bishop, J. A. *et al.* Melanocytic nevi, nevus genes, and melanoma risk in a large case-control study in the United Kingdom. *Cancer Epidemiol. Biomarkers Prev.* **19**, 2043–2054 (2010).
40. Pooley, K. A. *et al.* Common single-nucleotide polymorphisms in DNA double-strand break repair genes and breast cancer risk. *Cancer Epidemiol. Biomarkers Prev.* **17**, 3482–3489 (2008).
41. Pooley, K. A. *et al.* No association between TERT-CLPTM1L single nucleotide polymorphism rs401681 and mean telomere length or cancer risk. *Cancer Epidemiol. Biomarkers Prev.* **19**, 1862–1865 (2010).
42. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
43. Pavay, S. *et al.* Microarray expression profiling in melanoma reveals a BRAF mutation signature. *Oncogene* **23**, 4060–4067 (2004).
44. Strauch, K. Parametric linkage analysis with automatic optimization of the disease model parameters. *Am. J. Hum. Genet.* **73** (Suppl. 1), A2624 (2003).
45. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, e45 (2001).
46. Oancea, E. *et al.* TRPM1 forms ion channels associated with melanin content in melanocytes. *Sci. Signal.* **2**, ra21 (2009).

Alternatively activated macrophages produce catecholamines to sustain adaptive thermogenesis

Khoa D. Nguyen^{1,2*}, Yifu Qiu^{2*}, Xiaojin Cui², Y. P. Sharon Goh^{1,2}, Julia Mwangi², Tovo David², Lata Mukundan², Frank Brombacher³, Richard M. Locksley^{4,5,6} & Ajay Chawla^{2,7}

All homeotherms use thermogenesis to maintain their core body temperature, ensuring that cellular functions and physiological processes can continue in cold environments^{1–3}. In the prevailing model of thermogenesis, when the hypothalamus senses cold temperatures it triggers sympathetic discharge, resulting in the release of noradrenaline in brown adipose tissue and white adipose tissue^{4,5}. Acting via the β_3 -adrenergic receptors, noradrenaline induces lipolysis in white adipocytes⁶, whereas it stimulates the expression of thermogenic genes, such as PPAR- γ coactivator 1a (*Ppargc1a*), uncoupling protein 1 (*Ucp1*) and acyl-CoA synthetase long-chain family member 1 (*Acs1l*), in brown adipocytes^{7–9}. However, the precise nature of all the cell types involved in this efferent loop is not well established. Here we report in mice an unexpected requirement for the interleukin-4 (IL-4)-stimulated program of alternative macrophage activation in adaptive thermogenesis. Exposure to cold temperature rapidly promoted alternative activation of adipose

tissue macrophages, which secrete catecholamines to induce thermogenic gene expression in brown adipose tissue and lipolysis in white adipose tissue. Absence of alternatively activated macrophages impaired metabolic adaptations to cold, whereas administration of IL-4 increased thermogenic gene expression, fatty acid mobilization and energy expenditure, all in a macrophage-dependent manner. Thus, we have discovered a role for alternatively activated macrophages in the orchestration of an important mammalian stress response, the response to cold.

Mice housed at the thermoneutral temperature of 30 °C do not require adaptive thermogenesis, whereas those housed in colder environments depend on brown adipose tissue (BAT) thermogenesis to maintain their body temperature¹⁰. Thus, to understand the relationship between temperature and macrophage activation, we profiled the status of BAT and white adipose tissue (WAT) macrophages in mice chronically housed at 30 °C (thermoneutrality), 22 °C (normal

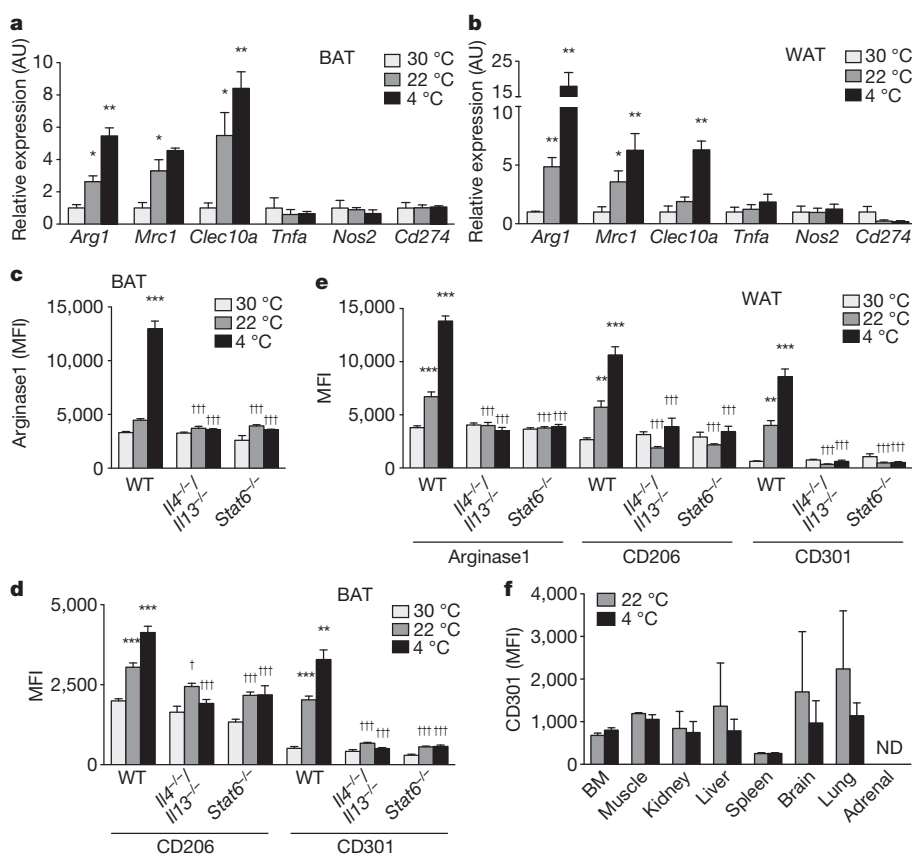


Figure 1 | Exposure to cold environment induces alternative activation of adipose tissue macrophages. **a, b**, Real-time PCR analysis of markers of alternative and classical activation in BAT (**a**) and WAT (**b**) of wild-type (WT) mice chronically housed at 30 °C, 22 °C, or acutely subjected to a 4 °C challenge from 22 °C ($n = 4$ per temperature). Expression of all genes is normalized to their relative expression at 30 °C in wild-type mice. **c–e**, Expression of alternative activation markers Arg1, CD206 and CD301 was monitored by flow cytometry in BAT (**c, d**) and WAT (**e**) macrophages of wild-type, *Il4*^{−/−}/*Il13*^{−/−} and *Stat6*^{−/−} mice housed at 30 °C, 22 °C and 4 °C ($n = 4–5$ per genotype and temperature). MFI, median fluorescence intensity. **f**, Alternative activation of tissue macrophages was monitored at 22 °C and 4 °C by quantifying expression of CD301. BM, bone marrow. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ comparison between wild-type mice at 30 °C and 22 °C, or between 22 °C and 4 °C. † $P < 0.05$, †† $P < 0.001$ comparison between wild-type and various knockout mice at the same temperature. All data are presented as mean \pm s.e.m.

¹Immunology Program, Stanford University, Palo Alto, California 94305, USA. ²Cardiovascular Research Institute, University of California, San Francisco, California 94158-9001, USA. ³Institute of Infectious Disease and Molecular Medicine, University of Cape Town, South Africa. ⁴Howard Hughes Medical Institute, University of California, San Francisco, California 94158-9001, USA. ⁵Departments of Medicine, University of California, San Francisco, California 94158-9001, USA. ⁶Microbiology & Immunology, University of California, San Francisco, California 94158-9001, USA. ⁷Departments of Physiology and Medicine, University of California, San Francisco, California 94158-9001, USA.

*These authors contributed equally to this work.

housing temperature), or after an acute challenge to 4 °C. Gene expression profiling revealed a progressive increase in the expression of alternative activation messenger RNAs^{11,12}, including *Arg1*, *Mrc1* and *Clec10a*, in BAT and WAT of mice exposed to colder temperatures (Fig. 1a, b). In contrast, expression of classical activation markers was unchanged by cold exposure (Fig. 1a, b). This correlation between alternative macrophage activation and exposure to colder environments was further verified using flow cytometry. In wild-type mice, exposure to progressively lower temperatures increased expression of CD206 (encoded by *Mrc1*), CD301 (*Clec10a*) and arginase 1 (*Arg1*) proteins in BAT and WAT macrophages (Fig. 1c–e and Supplementary Fig. 1a–f). Notably, disruption of IL-4/IL-13 signalling, as in *Il4/Il13*^{-/-} and *Stat6*^{-/-} mice¹³, completely abrogated the

cold-induced increase in alternative activation of BAT and WAT macrophages, as assessed by expression of CD206, CD301 and *Arg1* (Fig. 1c–e and Supplementary Fig. 1a–f). This was a specific defect in cold-induced alternative activation because loss of IL-4/IL-13 signalling did not introduce a classical activation bias in BAT and WAT macrophages (Supplementary Fig. 2a–d). Finally, acute exposure of mice to 4 °C failed to induce alternative macrophage activation in other tissues, including skeletal muscle and liver (Fig. 1f), indicating that BAT and WAT alternative activation is an adaptive response for acclimation to cold.

To investigate the importance of alternative macrophage activation in cold-induced thermogenesis, we challenged mice lacking alternatively activated macrophages to cold temperatures. Unlike wild-type

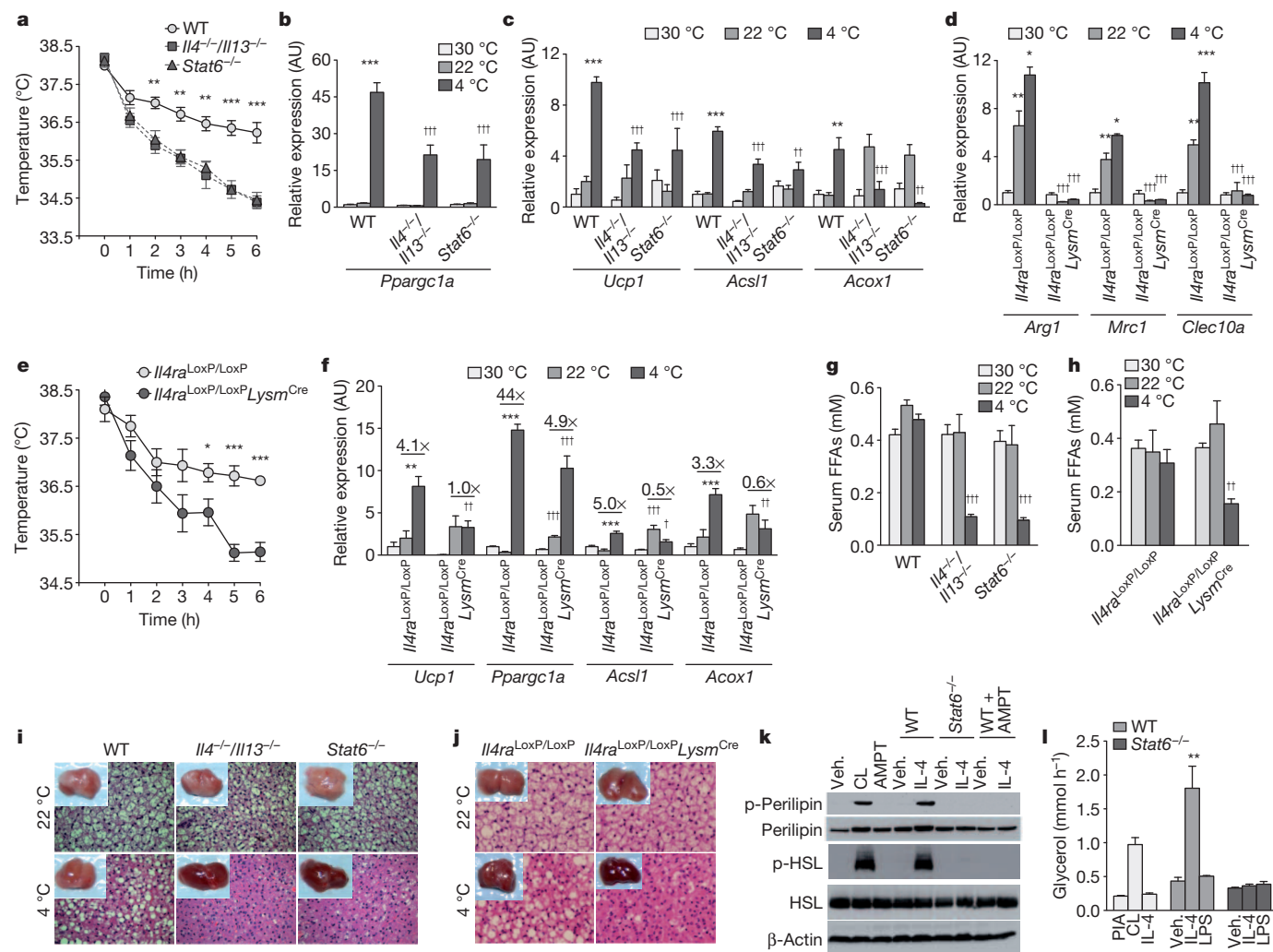


Figure 2 | Cold-induced metabolic adaptations require alternatively activated macrophages. **a**, Core body temperature of wild-type, *Il4^{-/-}/Il13^{-/-}* and *Stat6^{-/-}* mice during a cold challenge at 4 °C ($n = 8$ per genotype and temperature). **b**, **c**, Real-time PCR analysis of thermogenic genes in BAT of wild-type, *Il4^{-/-}/Il13^{-/-}* and *Stat6^{-/-}* mice housed at 30 °C, 22 °C or subjected to 4 °C cold challenge ($n = 4$ –5 per genotype and temperature). Expression of all genes is normalized to their relative expression at 30 °C in wild-type mice. **d**, Expression of alternatively activated mRNAs in BAT of *Il4ra^{LoxP/LoxP}* and *Il4ra^{LoxP/LoxP} Lysm^{Cre}* mice housed at various temperatures ($n = 5$ per genotype and temperature). **e**, Core body temperature of *Il4ra^{LoxP/LoxP}* and *Il4ra^{LoxP/LoxP} Lysm^{Cre}* mice during exposure to 4 °C ($n = 5$ –6 per genotype and temperature). **f**, BAT of *Il4ra^{LoxP/LoxP}* and *Il4ra^{LoxP/LoxP} Lysm^{Cre}* mice was analysed by real-time PCR for expression of thermogenic and β -oxidation genes ($n = 5$ per genotype and temperature). Expression of all genes is normalized to their relative expression at 30 °C in *Il4ra^{LoxP/LoxP}* mice. **g**, Serum free fatty acid (FFA) levels in wild-type, *Il4^{-/-}/Il13^{-/-}* and *Stat6^{-/-}* mice housed at 30 °C, 22 °C and 4 °C ($n = 5$ –8 per

genotype). **h**, Serum FFAs in *Il4ra^{LoxP/LoxP}* and *Il4ra^{LoxP/LoxP} Lysm^{Cre}* mice housed at the three temperatures ($n = 5$ –11 per genotype). **i**, Representative gross and microscopic (haematoxylin and eosin staining) histology of BAT from wild-type, *Il4^{-/-}/Il13^{-/-}* and *Stat6^{-/-}* mice at 22 °C and after exposure to 4 °C for 6 h. **j**, Representative gross and microscopic (haematoxylin and eosin staining) histology of BAT from *Il4ra^{LoxP/LoxP}* and *Il4ra^{LoxP/LoxP} Lysm^{Cre}* mice at 22 °C and after 6 h exposure to 4 °C. **k**, Immunoblot analysis for serine-phosphorylated perilipin, total perilipin, serine-phosphorylated HSL and total HSL in 3T3-L1 adipocytes treated with PIA, CL-316243, IL-4 or macrophage conditioned medium (with/without IL-4 and AMPT) for 15 min. **l**, Glycerol release by 3T3-L1 adipocytes after 6-h treatment with PIA, CL-316243, IL-4 or macrophage conditioned medium ($n = 5$ –7). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ compared to comparison between wild-type or *Il4ra^{LoxP/LoxP}* mice at 30 °C and those at 22 °C, or at 22 °C and 4 °C. † $P < 0.05$, †† $P < 0.01$, ††† $P < 0.001$ comparison between knockouts and wild-type or *Il4ra^{LoxP/LoxP}* mice at the same temperature. All data are presented as mean \pm s.e.m.

mice, $Il4^{-/-}/Il13^{-/-}$ and $Stat6^{-/-}$ mice showed a drop in core body temperature when exposed to temperatures of 4 °C (Fig. 2a). In wild-type mice, to counteract the change in environmental temperature, thermogenic genes (*Ppargc1a* and *Ucp1*) and the β -oxidation genes (*Acox1* and *Acs1l*) were induced in BAT. This induction of thermogenic genes was blunted in BAT of $Il4^{-/-}/Il13^{-/-}$ and $Stat6^{-/-}$ mice (Fig. 2b, c). To determine whether the observed defects in cold-induced thermogenesis were a direct consequence of the loss of alternatively activated macrophages, we disrupted IL-4/IL-13 signalling in myeloid cells by breeding conditional $Il4^{LoxP/LoxP}$ with $Lysm^{Cre}$ mice¹⁴. BAT macrophages in $Il4^{LoxP/LoxP}Lysm^{Cre}$ mice displayed impairment in alternative activation at 22 °C and 4 °C (Fig. 2d), which was sufficient to render mutant mice susceptible to cold-induced hypothermia (Fig. 2e). $Il4^{LoxP/LoxP}Lysm^{Cre}$ mice also showed defects in expression of cold-inducible thermogenic genes, including *Ucp1*, *Acox1*, *Acs1l* and *Ppargc1a* (Fig. 2f). Comparable results were obtained in a second model when macrophages were pharmacologically depleted in BAT using clodronate-containing liposomes (Supplementary Fig. 3a–e), which selectively deplete tissue macrophages and circulating monocytes but not neutrophils (Supplementary Fig. 4a, b). Moreover, expression of skeletal muscle mitochondrial genes implicated in thermogenesis was unaltered (Supplementary Fig. 5a), indicating a primary defect in non-shivering thermogenesis. Serum triglyceride levels and expression of lipogenic genes in liver were similarly unchanged across the genotypes and temperatures (Supplementary Table 1 and Supplementary Fig. 5b). Finally, defects in cold-induced thermogenesis were also observed in $Stat6^{-/-}$ mice on the C57BL/6J background (Supplementary Fig. 5c, d).

During cold exposure, β -adrenergic signalling in white adipocytes stimulates the release of free fatty acids to fuel uncoupled respiration in BAT¹⁶. Because WAT macrophages also undergo alternative activation upon cold challenge (Fig. 1b, e), we examined whether a defect in alternative macrophage activation was associated with impaired release of free fatty acids. Indeed, compared to wild-type mice, circulating levels of free fatty acids were reduced by ~75% in $Il4/Il13^{-/-}$ and $Stat6^{-/-}$ mice (Fig. 2g). Serum free fatty acid levels were similarly

reduced by ~65% in $Il4^{LoxP/LoxP}Lysm^{Cre}$ mice at 4 °C (Fig. 2h). Consistent with reduced release of fatty acids, gross and microscopic histology revealed that all mutant mice impaired in alternative macrophage activation had exhausted their lipid stores in BAT (Fig. 2i, j). Correspondingly, mice deficient in IL-4/IL-13 signalling or alternatively activated macrophages lost less weight during the cold challenge (Supplementary Table 1).

To explore whether factors released by alternatively activated macrophages work in *trans* to stimulate lipolysis of stored triglycerides, we used differentiated 3T3-L1 cells to study triglyceride lipolysis *in vitro*. Treatment of adipocytes with conditioned medium from alternatively activated macrophages induced phosphorylation of perilipin and hormone sensitive lipase (HSL), lipases that are phosphorylated by protein kinase A in response to adrenergic signalling (Fig. 2k). The phosphorylation of perilipin A releases CGI-58, allowing it to interact with Pnpla2 to enhance the lipolysis of stored triglycerides^{16,17}. Indeed, paralleling the increase in perilipin phosphorylation, triglyceride lipolysis, as quantified by glycerol release, was increased by ~4.5-fold in adipocytes treated with conditioned medium from alternatively activated macrophages (Fig. 2l). No significant increase in phosphorylation of perilipin, HSL, or triglyceride lipolysis was observed when adipocytes were exposed to conditioned medium from $Stat6^{-/-}$ macrophages (Fig. 2k, l). Together, these data indicate that alternatively activated macrophages coordinate the thermogenic response during cold exposure by increasing the thermogenic capacity of BAT and mobilizing fatty acids to fuel uncoupled respiration.

The requirement for alternatively activated macrophages in fatty acid mobilization and thermogenic gene induction prompted us to investigate whether WAT and BAT macrophages might be an important source of catecholamines. In this regard, catecholamine production by classically activated macrophages has previously been shown to promote inflammation-induced injury^{18,19}. Intracellular staining for tyrosine hydroxylase (Th), dopa decarboxylase (Ddc) and dopamine β -hydroxylase (Dbh) revealed that all three catecholamine-synthesizing enzymes were induced in macrophages upon stimulation with IL-4 (Supplementary Fig. 6a–f). This induction was a bona fide part of

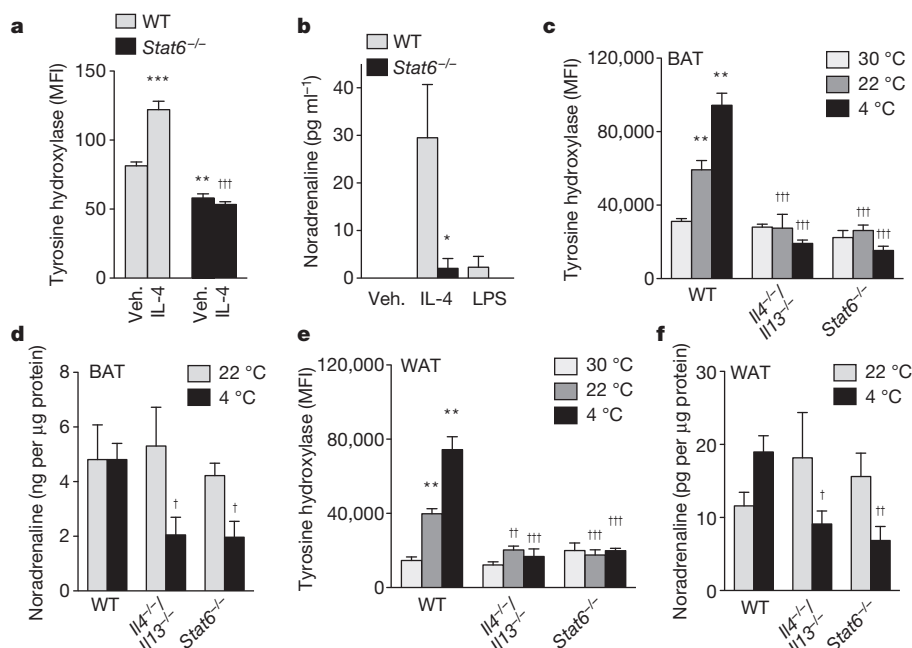


Figure 3 | Alternately activated macrophages produce catecholamines.

a, Expression of tyrosine hydroxylase in wild-type and $Stat6^{-/-}$ peritoneal macrophages treated with vehicle (Veh.) or IL-4 ($n = 5$ per genotype and condition). **b**, Noradrenaline secretion by wild-type and $Stat6^{-/-}$ bone-marrow-derived macrophages stimulated with IL-4 or LPS ($n = 5$ per genotype and condition). **c**, **e**, Tyrosine hydroxylase expression in BAT (**c**) and WAT

(**e**) macrophages of wild-type and $Stat6^{-/-}$ mice at 30 °C, 22 °C and 4 °C ($n = 5$ per genotype and temperature). **d**, **f**, Noradrenaline content of BAT (**d**) and WAT (**f**) at 22 °C and 4 °C of wild-type and $Stat6^{-/-}$ mice ($n = 4-5$ per genotype and temperature). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ compared to wild type. † $P < 0.05$, †† $P < 0.01$, ††† $P < 0.001$ compared to wild type with IL-4 at 4 °C samples. All data are presented as mean \pm s.e.m.

alternative activation because IL-4 failed to induce Th, the rate-limiting step in the synthesis of catecholamines²⁰, in macrophages lacking STAT6 (Fig. 3a and Supplementary Fig. 6g). Congruent with this, stimulation of macrophages with IL-4, but not lipopolysaccharide (LPS), increased secretion of noradrenaline and adrenaline into the culture medium in a STAT6-dependent manner (Fig. 3b and Supplementary Fig. 7a). Furthermore, treatment of wild-type macrophages with α -methyltyrosine, a specific inhibitor of tyrosine hydroxylase¹⁸, inhibited secretion of noradrenaline into the culture medium and abrogated its lipolytic activity on cultured adipocytes (Fig. 2k and Supplementary Fig. 7b, c).

Next, we examined catecholamine synthesis by adipose tissue macrophages. At thermoneutrality (30 °C), expression of Th in BAT and WAT macrophages was the lowest (Fig. 3c, e). The expression progressively increased as mice were exposed to colder temperatures (Fig. 3c, e), and was restricted to Ly6C^{lo-mid}CD301⁺ alternatively activated BAT and WAT macrophages (Supplementary Fig. 8a, b). Consistent with this, loss of IL-4/IL-13 signalling abrogated cold-induced expression of Th in BAT and WAT macrophages (Fig. 3c, e) and reduced noradrenaline content of these adipose tissues by ~50–60% in *Il4*^{-/-}/*Il13*^{-/-} and *Stat6*^{-/-} mice (Fig. 3d, f and Supplementary Table 2). This decrease in BAT and WAT catecholamine content was a direct consequence of

loss of alternative activation because similar changes were observed in *Il4*^{LoxP/LoxP}*Lysm*^{Cre} mice. Specifically, cold exposure failed to induce Th protein in BAT and WAT macrophages (Supplementary Fig. 9a, c), resulting in a 70–80% reduction of noradrenaline content in BAT and WAT of *Il4*^{LoxP/LoxP}*Lysm*^{Cre} mice (Supplementary Fig. 9b, d). In contrast, catecholamine content of serum and other tissues was unaffected (Supplementary Table 3). Finally, IL-4 stimulation of primary human monocytes or U937 cells enhanced their alternative activation, catecholamine production and lipolytic activity on adipocytes, suggesting conservation of these pathways across species (Supplementary Fig. 10a–h).

These data prompted us to investigate whether the β_3 -adrenergic agonist CL-316243 can rescue the thermogenic defect in *Il4*^{-/-}/*Il13*^{-/-} mice²¹. Indeed, a single injection of CL-316243 increased core body temperature and thermogenic gene expression in *Il4*^{-/-}/*Il13*^{-/-} mice (Supplementary Fig. 11a, b). The restoration of core body temperature by CL-316243 also normalized weight loss and BAT histology in *Il4*^{-/-}/*Il13*^{-/-} mice housed at 4 °C (Supplementary Fig. 11c–e), including the reappearance of lipid droplets in brown adipocytes. The increased accumulation of lipid droplets probably resulted from enhanced mobilization of free fatty acids and induction of lipogenic genes, such as *Lpl*, *Hmgcs1* and *Dgat1*, in BAT of *Il4*^{-/-}/*Il13*^{-/-} mice

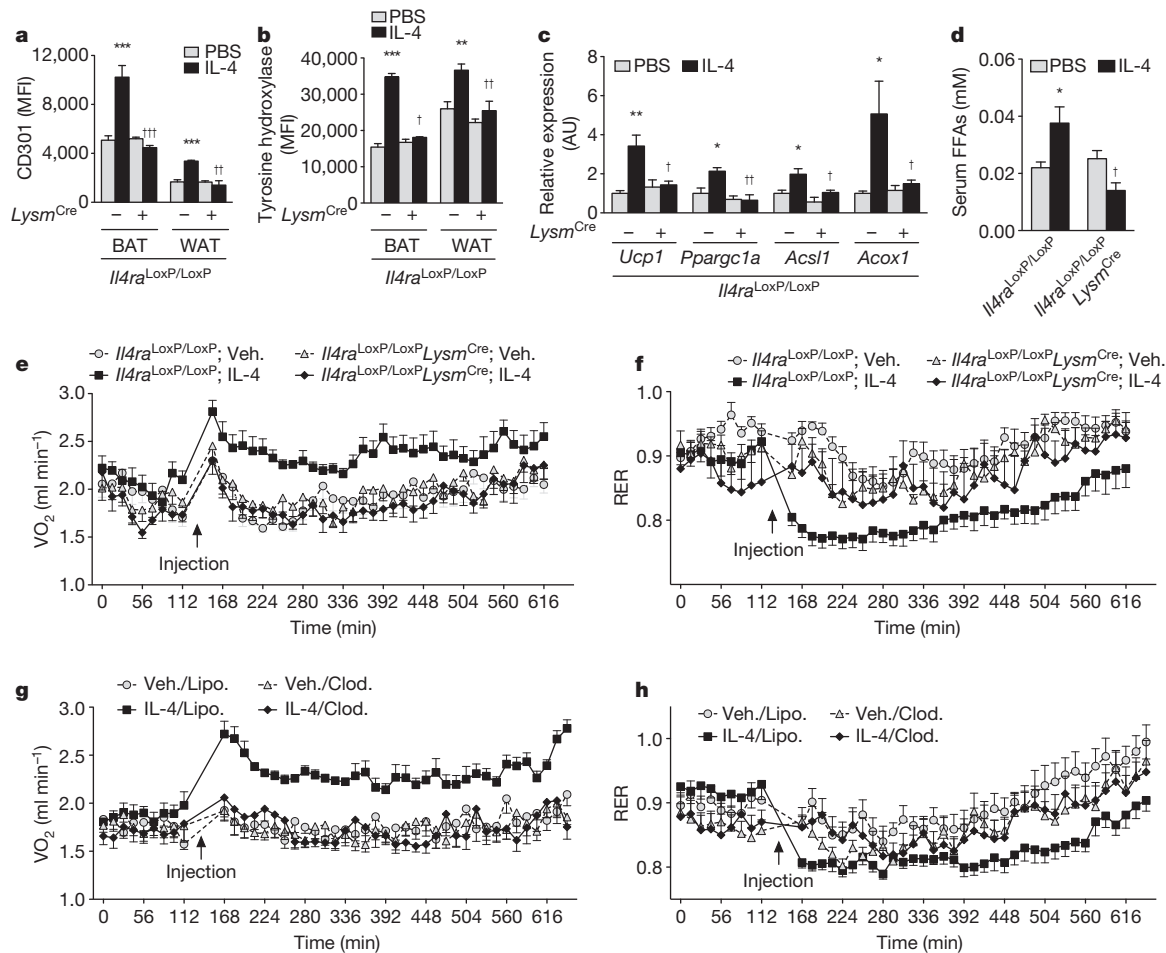


Figure 4 | Alternative activation of macrophages increases energy expenditure. **a, b**, Expression of alternative activation marker CD301 (**a**) and Th (**b**) in adipose tissue macrophages from *Il4*^{LoxP/LoxP} and *Il4*^{LoxP/LoxP}*Lysm*^{Cre} mice treated with vehicle (Veh.) or IL-4 for 6 h at 22 °C ($n = 4-5$ per genotype and condition). **c**, Real-time PCR for thermogenic genes in BAT of *Il4*^{LoxP/LoxP} and *Il4*^{LoxP/LoxP}*Lysm*^{Cre} mice treated with Veh. or IL-4 for 6 h at 22 °C ($n = 4-5$ per genotype and condition). **d**, Serum free fatty acid (FFA) levels in *Il4*^{LoxP/LoxP} and *Il4*^{LoxP/LoxP}*Lysm*^{Cre} mice treated with Veh. or IL-4 for 6 h at 22 °C ($n = 4-5$ per genotype and condition). **e, f**, Quantification of energy expenditure in *Il4*^{LoxP/LoxP} and

Il4^{LoxP/LoxP}*Lysm*^{Cre} mice treated with vehicle (Veh.) or IL-4 ($n = 7-9$ per genotype and condition). **e, f**, Oxygen consumption (**e**; VO₂) and respiratory exchange ratio (**f**; RER). **g, h**, Quantification of energy expenditure in wild-type mice after macrophage depletion ($n = 8$ per condition). Mice were injected with empty liposomes (Lipo.) or clodronate-containing liposomes (Clod.) 24 h before energy expenditure studies. All data were collected during the light cycle. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ compared to *Il4*^{LoxP/LoxP} with Veh. † $P < 0.05$, †† $P < 0.01$, ††† $P < 0.001$ compared to *Il4*^{LoxP/LoxP} with IL-4. All data are presented as mean \pm s.e.m.

treated with CL-316243 (Supplementary Figs 11f, g and 12a, b). Hence, alternatively activated macrophages are an unexpected source of noradrenaline that sustains the metabolic adaptation to cold.

A hallmark of cold-induced thermogenesis is an increase in uncoupled respiration and energy expenditure by noradrenaline¹⁰. Because we observed that IL-4 driven alternatively activated macrophages release noradrenaline in BAT and WAT in response to cold, we next examined the metabolic effects of IL-4 in wild-type mice. Injection of IL-4 induced alternative activation and Th expression in BAT and WAT macrophages (Supplementary Fig. 13a, b). As expected, the strongest effects of IL-4 were observed at thermoneutrality, when basal alternative activation and Th expression was lowest. However, administration of IL-4 was sufficient to augment alternative activation and Th expression in mice housed at 22 °C and 4 °C (Supplementary Fig. 13a, b). Concomitant with the induction of alternative activation, noradrenaline content and thermogenic gene expression in BAT, as well as fatty acid levels in serum, increased after administration of IL-4 (Supplementary Fig. 13c–e). Finally, administration of IL-4 enhanced expression of Th in alternatively activated macrophages taking residence in other tissues, including liver, spleen, lung and bone marrow (Supplementary Figs 14a–f and 15a, b), albeit to a much lower degree than Th expression in BAT and WAT macrophages.

We next investigated whether acute administration of IL-4 to adapted animals could enhance oxygen consumption in a macrophage-dependent manner. As shown in Fig. 4a, administration of IL-4 promoted alternative activation of BAT and WAT macrophages in *Il4ra*^{LoxP/LoxP} mice but not *Il4ra*^{LoxP/LoxP}*Lysm*^{Cre} mice. This was accompanied by an increase in expression of Th in BAT and WAT macrophages, resulting in induction of thermogenic genes and release of free fatty acids (Fig. 4b–d). Furthermore, quantification of energy expenditure revealed that injection of IL-4 rapidly increased oxygen consumption in *Il4ra*^{LoxP/LoxP} but not *Il4ra*^{LoxP/LoxP}*Lysm*^{Cre} mice (Fig. 4e). Importantly, consistent with a shift from carbohydrate to fatty acid metabolism, administration of IL-4 decreased the respiratory exchange ratio (RER) in *Il4ra*^{LoxP/LoxP} mice (Fig. 4f). These changes in energy expenditure were independent of alterations in locomotor activity (Supplementary Fig. 15c). Furthermore, in wild-type mice, the stimulatory effect of IL-4 on energy expenditure showed a marked dependence on macrophages, as IL-4 failed to raise oxygen consumption or decrease RER in mice treated with clodronate-containing liposomes (Fig. 4g, h and Supplementary Fig. 15d). These findings provide direct evidence that actions of alternatively activated macrophages in BAT and WAT orchestrate the metabolic programs that constitute adaptive thermogenesis.

The data presented here show that alternatively activated macrophages participate *in vivo* in the regulation of adaptive and facultative aspects of non-shivering thermogenesis. In a macrophage-dependent manner, the administration of IL-4 raises energy expenditure in a facultative manner, whereas adaptation to lower temperatures is associated with polarization of BAT and WAT macrophages to the alternative state. Moreover, the secretion of noradrenaline by alternatively activated macrophages allows these cells to coordinate the thermogenic response in animals experiencing cold stress. Thus, we propose that, in addition to the sympathetic nerves, cells of the haematopoietic system, such as alternatively activated macrophages, constitute a second, parallel circuit for controlling non-shivering thermogenesis.

METHODS SUMMARY

Male mice, 8–12 weeks old, were used in all experiments. Breeding pairs of wild-type and *Stat6*^{−/−} mice on a BALB/c background were purchased from the Jackson Laboratory, and *Il4*^{−/−}*Il13*^{−/−}, *Il4ra*^{LoxP/LoxP} and *Lysm*^{Cre} mice on the BALB/c background were obtained from the Locksley or Brombacher laboratories. For cold challenge experiments, mice were fed *ad libitum* and individually housed in cages that had been pre-chilled at 4 °C (ref. 8). Core body temperature was monitored hourly by a rectal temperature probe (Physitemp). For the thermoneutrality experiments, mice were adapted to 30 °C in a laboratory incubator (Darwin Chambers) for 2–4 weeks before experimentation. For rescue experiments, the β_3 -adrenergic agonist CL-316243 (Sigma) was injected intraperitoneally at 0.1 mg kg^{−1} 30 min

before the cold challenge. Tissues were collected at the end of a 6-h cold challenge, and processed for RNA and protein analyses. To deplete macrophages, mice were injected intraperitoneally with two doses of clodronate-containing or empty liposomes (400 μ l and 100 μ l at 24 h and 30 min, respectively, before initiation of experiment)²². Depletion was confirmed by flow cytometric analysis of monocytes and macrophages in blood, adipose tissues and spleen. Cohorts of ≥ 4 mice per genotype or treatment were assembled for all *in vivo* studies, which were repeated 2–3 independent times. All data are presented as mean \pm s.e.m.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 6 April; accepted 18 October 2011.

Published online 20 November 2011.

1. Cannon, B. & Nedergaard, J. Brown adipose tissue: function and physiological significance. *Physiol. Rev.* **84**, 277–359 (2004).
2. Lowell, B. B. & Spiegelman, B. M. Towards a molecular understanding of adaptive thermogenesis. *Nature* **404**, 652–660 (2000).
3. Tseng, Y. H., Cypess, A. M. & Kahn, C. R. Cellular bioenergetics as a target for obesity therapy. *Nature Rev. Drug Discov.* **9**, 465–482 (2010).
4. Nakamura, K. & Morrison, S. F. A thermosensory pathway that controls body temperature. *Nature Neurosci.* **11**, 62–71 (2008).
5. Morrison, S. F., Nakamura, K. & Madden, C. J. Central control of thermogenesis in mammals. *Exp. Physiol.* **93**, 773–797 (2008).
6. Nedergaard, J., Bengtsson, T. & Cannon, B. New powers of brown fat: fighting the metabolic syndrome. *Cell Metab.* **13**, 238–240 (2011).
7. Ellis, J. M. *et al.* Adipose acyl-CoA synthetase-1 directs fatty acids toward β -oxidation and is required for cold thermogenesis. *Cell Metab.* **12**, 53–64 (2010).
8. Enerbäck, S. *et al.* Mice lacking mitochondrial uncoupling protein are cold-sensitive but not obese. *Nature* **387**, 90–94 (1997).
9. Puigserver, P. *et al.* A cold-inducible coactivator of nuclear receptors linked to adaptive thermogenesis. *Cell* **92**, 829–839 (1998).
10. Cannon, B. & Nedergaard, J. Nonshivering thermogenesis and its adequate measurement in metabolic studies. *J. Exp. Biol.* **214**, 242–253 (2011).
11. Gordon, S. Alternative activation of macrophages. *Nature Rev. Immunol.* **3**, 23–35 (2003).
12. Odegaard, J. I. & Chawla, A. Alternative macrophage activation and metabolism. *Annu. Rev. Pathol.* **6**, 275–297 (2011).
13. Martinez, F. O., Helming, L. & Gordon, S. Alternative activation of macrophages: an immunologic functional perspective. *Annu. Rev. Immunol.* **27**, 451–483 (2009).
14. Herbert, D. R. *et al.* Alternative macrophage activation is essential for survival during schistosomiasis and downmodulates T helper 1 responses and immunopathology. *Immunity* **20**, 623–635 (2004).
15. Watt, M. J. *et al.* Reduced plasma FFA availability increases net triacylglycerol degradation, but not GPAT or HSL activity, in human skeletal muscle. *Am. J. Physiol. Endocrinol. Metab.* **287**, E120–E127 (2004).
16. Haemmerle, G. *et al.* Defective lipolysis and altered energy metabolism in mice lacking adipose triglyceride lipase. *Science* **312**, 734–737 (2006).
17. Lass, A. *et al.* Adipose triglyceride lipase-mediated lipolysis of cellular fat stores is activated by CGI-58 and defective in Chanarin-Dorfman Syndrome. *Cell Metab.* **3**, 309–319 (2006).
18. Flierl, M. A. *et al.* Phagocyte-derived catecholamines enhance acute inflammatory injury. *Nature* **449**, 721–725 (2007).
19. Brown, S. W. *et al.* Catecholamines in a macrophage cell line. *J. Neuroimmunol.* **135**, 47–55 (2003).
20. Zhou, Q. Y., Quaife, C. J. & Palmiter, R. D. Targeted disruption of the tyrosine hydroxylase gene reveals that catecholamines are required for mouse fetal development. *Nature* **374**, 640–643 (1995).
21. Yoshida, T., Sakane, N., Wakabayashi, Y., Umekawa, T. & Kondo, M. Anti-obesity and anti-diabetic effects of CL 316,243, a highly specific β_3 -adrenoceptor agonist, in yellow KK mice. *Life Sci.* **54**, 491–498 (1994).
22. Kostel, A. *et al.* Weight loss and lipolysis promote a dynamic immune response in murine adipose tissue. *J. Clin. Invest.* **120**, 3466–3479 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank members of the Chawla laboratory, A. Loh and C.-H. Lee for comments on the manuscript, and F. Kraemer for guidance on *in vitro* lipolysis assays. This work was supported by grants from the NIH (DK076760, HL076746, DK094641), Larry L. Hillblom Foundation Network Grant and an NIH Director's Pioneer Award (DP1OD006415) to A.C. Support was provided by Stanford Graduate Fellowship (K.D.N.) and A-STAR Fellowship (Y.P.G.). All animal care was in accordance with Stanford University's A-PLAC and UCSF's IACUC guidelines.

Author Contributions K.D.N. and Y.Q. performed the experiments with assistance from X.C., J.M., T.D., Y.P.S.G. and L.M.; F.B., R.M.L. and A.C. were involved in project planning; K.D.N., Y.Q. and A.C. designed the experiments, analysed the data and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.C. (ajay.chawla@ucsf.edu).

METHODS

Animals and *in vivo* studies. Male mice, 8–12 weeks old, were used in all experiments. Breeding pairs of wild-type and *Stat6*^{−/−} mice on a BALB/cJ background were purchased from the Jackson Laboratory, and *Il4*^{−/−}/*Il13*^{−/−}, *Il4ra*^{LoxP/LoxP} and *Lysm*^{Cre} mice on the BALB/cJ background were obtained from the Locksley or Brombacher laboratories. For cold challenge experiments, mice were fed *ad libitum* and individually housed in cages that had been pre-chilled at 4 °C (ref. 8). Core body temperature was monitored hourly by a rectal temperature probe (Physitemp). For the thermoneutrality experiments, mice were adapted to 30 °C in a laboratory incubator (Darwin Chambers) for 2–4 weeks before experimentation. For rescue experiments, the β_3 -adrenergic agonist CL-316243 (Sigma) was injected intraperitoneally at 0.1 mg kg^{−1} 30 min before the cold challenge. Tissues were collected at the end of a 6-h cold challenge, and processed for RNA and protein analyses. To deplete macrophages, mice were injected intraperitoneally with two doses of clodronate-containing or empty liposomes (400 μ l and 100 μ l at 24 h and 30 min, respectively, before initiation of experiment)²². Depletion was confirmed by flow cytometric analysis of monocytes and macrophages in blood, adipose tissues and spleen. Cohorts of ≥ 4 mice per genotype or treatment were assembled for all *in vivo* studies, which were repeated 2–3 independent times.

Flow cytometry and immunoblot analysis. Adipose tissues were minced and digested with collagenase I (3 mg ml^{−1}, Worthington) for 45 min at 37 °C in a shaker (400 r.p.m.). The digested cell suspension was centrifuged at 1,600 r.p.m. for 5 min to separate stromal-vascular fraction from adipocytes. Pelleted cells were re-suspended in FACS buffer (PBS containing 5% FBS and 1% L-glutamine) and passed through a 40 μ m strainer (BD Biosciences) to remove large cellular debris. Antibodies directed against mouse CD3, B220, Ly6G, CD45, CD49b, CD11c and F4/80 (Biolegend), Siglec F (BD Biosciences), Fc ϵ R1 (eBioscience), tyrosine hydroxylase (Origene), Ly6C, dopamine β -hydroxylase and dopa decarboxylase (Abcam), arginase-1 (Santa Cruz Biotech), and anti-rabbit and anti-mouse IgG (Invitrogen) were used for flow cytometric analysis. Samples were stored in FACS buffer with 1% paraformaldehyde at 4 °C before analysis. Data was acquired on LSRII (BD Biosciences) and data analysis was performed using FlowJo (Treestar). For analysis of mitochondrial proteins in muscle, soleus muscle from mice housed at 22 °C and 4 °C was lysed using TissueLyser II (Qiagen), and antibodies directed against Cox1 (Invitrogen) or Cpt1b (Alpha Diagnostic International) were used to detect mitochondrial proteins.

Macrophage culture and stimulation. Bone-marrow-derived macrophages (BMDMs) were cultured as previously described²³. Classical or alternative activation was induced in BMDMs by stimulation with LPS (10 ng ml^{−1}) or IL-4 (10 ng ml^{−1}) for 24 h, respectively. Macrophages were elicited into the peritoneal cavity by injection of thioglycollate (3 ml, BD Biosciences). To promote alternative activation of elicited macrophages, mice were given a single injection of IL-4 (2 μ g) complexed to anti-IL-4 antibody (10 μ g) 3 days after injection of thioglycollate, and elicited macrophages were recovered 24 h later. After washing two times, elicited macrophages were cultured in low glucose (LG)-DMEM with 3% BSA and macrophage-conditioned medium was collected 24 h later. Human monocytes, which were isolated from whole blood by magnetic purification with CD14 microbeads (Miltenyi Biotec), and the human macrophage cell line U937 were cultured in RPMI with 5% FBS and 1% L-glutamine and stimulated with LPS (10 ng ml^{−1}) or IL-4 (10 ng ml^{−1}) for 24 h to induce classical or alternative activation, respectively. Conditioned media was collected and supplemented with 3% BSA for lipolysis assays. In some experiments, α -methyl-p-tyrosine (AMPT, 2 mM, Sigma) was added to cultured macrophages to inhibit tyrosine hydroxylase.

Quantitative RT-PCR. Tissues were homogenized in Trizol (Invitrogen) and total RNA was isolated using the RNeasy kit (Qiagen) and used as template for cDNA synthesis (Origene). Quantitative PCR reactions were carried out in triplicate using the CFX384 real-time PCR detection system (Bio-Rad). Relative expression level of mRNAs was calculated by the comparative threshold cycle method using 36B4 as an internal control²⁴. The following primers were used in these studies: *Il1* forward 5'-GAAGAAGAGCCCATCTCTG-3', reverse 5'-TCATCTCGGAGCCTGTAGTG-3'; *Il6* forward 5'-AGTCCGGAGAGGA GACTTCA-3', reverse 5'-TTGCCATTGCACAACTCTTT-3'; *Acs1l* forward

5'-TGGGGTGGAAATCATCAGCC-3', reverse 5'-CACAGCATTACACACTG TACAACGG-3'; *Acox* forward 5'-GGTGGACCTCTGTCTTGTTC-3', reverse 5'-AAACCTTCAGGCCCAAGTGAG-3'; *Ucp1* forward 5'-GTGAAGGTCA GAATGCAAGC-3', reverse 5'-AGGGCCCCCTTCATGAGGTC-3'; *Ppargc1a* forward 5'-CAACATGCTCAAGCCAAACCAACA-3', reverse 5'-CGCTCAAT AGTCTTGTCTCAAATGGG-3'; *Tnf* forward 5'-CCAAGGCGCCACATC TCCCT-3', reverse 5'-GCTTTCTGTGCTCATGGTGT-3'; *Nos2* forward 5'-ACCTTGGTGAAGGGACTGAG-3', reverse 5'-TCCGTTCTCTTCAGATT GAC-3'; *Arg1* forward 5'-AGACCACAGTCTGGCAGTTG-3', reverse 5'-CCACCCAAATGACACATAGG-3'; *Mrc1* forward 5'-TGATTACGAGCAG TGGAAGC-3', reverse 5'-GTTCCACCGTAAGCCCAATT-3'; *Clec10a* forward 5'-CTCTGGAGAGCAGTGGAG-3', reverse 5'-ACTTCCGAGCCGTTGT TCT-3'; *36B4* forward 5'-GAGACTGAGTACACCTTCCAC-3', reverse 5'-ATGCAGATGGATCAGCCAGG-3'.

Catecholamines and lipids. Catecholamines (Rocky Mountain Diagnostics), free fatty acids (Biovision) and glycerol (Abcam) were quantified in duplicate as per the manufacturers' protocols. For catecholamine ELISAs, tissues were homogenized by sonification in homogenization buffer (1 N HCl, 0.25 M EDTA, 1 M Na₂S₂O₅), and cellular debris was pelleted by centrifugation at 13,000 r.p.m. for 15 min at 4 °C. The cleared homogenates were collected and stored at −80 °C before quantification. All samples were normalized to total tissue protein concentration.

Adipocyte differentiation and lipolysis. The 3T3-L1 pre-adipocytes were grown in high glucose Dulbecco's modified Eagle's medium (HG-DMEM) supplemented with BCS (10%). Two days after confluence, differentiation was induced with insulin (10 μ g ml^{−1}), dexamethasone (1 μ M), and 3-isobutyl-1-methylxanthine (0.5 mM) in HG-DMEM containing FBS (10%). All subsequent media changes (every 2 days) were performed using HG-DMEM supplemented with FBS (10%) and insulin (10 μ g ml^{−1}). For the lipolysis studies, differentiated adipocytes were cultured in low glucose (LG)-DMEM supplemented with BSA (3%) for 16 h before stimulation with vehicle, CL-316243 (1 μ M), IL-4 (10 ng ml^{−1}) or macrophage conditioned medium for 15 min or 6 h to quantify phospho-HSL, phospho-perilipin or glycerol release, respectively. Basal lipolysis was quantified in the presence of N6-phenylisopropyl adenosine (PIA, 1 μ M). Glycerol release into the culture medium was quantified using the Free Glycerol Assay Kit (Abcam). For immunoblot analysis, treated adipocytes were lysed in lysis buffer (20 mM Tris-HCl, pH 7.5, 100 mM KCl, 0.1% Nonidet P-40, 1 mM EDTA, and 10% glycerol containing 1 mM phenylmethylsulphonyl fluoride, 1% protease inhibitor cocktail and 1% phosphatase inhibitor cocktails I/II) for 30 min at 4 °C. Total cellular protein extracts were separated on SDS-polyacrylamide gel electrophoresis, transferred to nitrocellulose membrane (Bio-Rad), and incubated with antibodies directed against HSL, serine-660 phosphorylated HSL (Cell Signaling), perilipin, or serine-552-phosphorylated perilipin (Vala Sciences). After incubation with the appropriate secondary antibodies, proteins were detected with SuperSignal West Pico Chemiluminescent Substrate (Thermo Scientific).

Energy expenditure. Oxygen consumption, RER and activity were quantified in 12-week-old male mice of various genotypes fed *ad libitum* using CLAMS (Columbus Instruments). Following acclimatization to CLAMS cages for 48 h, mice were given an intraperitoneal injection of recombinant IL-4 (45.5 μ g kg^{−1} body weight) at 11:00. Consumption rates of O₂ (VO₂) and release of CO₂ (VCO₂) were monitored for ~8 h every 14 min. Locomotor activity—the number of x-axis beam breaks—was monitored every minute. Data was collected during light cycle (9:00 to approximately 19:00).

Statistical analysis. All data are presented as mean \pm s.e.m. and analysed using Prism (Graphpad). Statistical significance was determined using the Student's *t*-test and two-way analysis of variance test. A *P* value of <0.05 was considered to be statistically significant, and is presented as * (*P* < 0.05), ** (*P* < 0.01), or *** (*P* < 0.001).

23. Odegaard, J. I. *et al.* Alternative M2 activation of Kupffer cells by PPAR δ ameliorates obesity-induced insulin resistance. *Cell Metab.* **7**, 496–507 (2008).
24. Odegaard, J. I. *et al.* Macrophage-specific PPAR γ controls alternative activation and improves insulin resistance. *Nature* **447**, 1116–1120 (2007).

Lyn is a redox sensor that mediates leukocyte wound attraction *in vivo*

Sa Kan Yoo¹, Taylor W. Starnes^{2*}, Qing Deng^{2*} & Anna Huttenlocher^{2,3}

Tissue wounding induces the rapid recruitment of leukocytes¹. Wounds and tumours—a type of ‘unhealed wound’²—generate hydrogen peroxide (H₂O₂) through an NADPH oxidase (NOX). This extracellular H₂O₂ mediates recruitment of leukocytes, particularly the first responders of innate immunity, neutrophils, to injured tissue^{3–6}. However, the sensor that neutrophils use to detect the redox state at wounds is unknown. Here we identify the Src family kinase (SFK) Lyn as a redox sensor that mediates initial neutrophil recruitment to wounds in zebrafish larvae. Lyn activation in neutrophils is dependent on wound-derived H₂O₂ after tissue injury, and inhibition of Lyn attenuates neutrophil wound recruitment. Inhibition of SFKs also disrupted H₂O₂-mediated chemotaxis of primary human neutrophils. *In vitro* analysis identified a single cysteine residue, C466, as being responsible for direct oxidation-mediated activation of Lyn. Furthermore, transgenic-tissue-specific reconstitution with wild-type Lyn and a cysteine mutant revealed that Lyn C466 is important for the neutrophil wound response and downstream signalling *in vivo*. This is the first identification, to our knowledge, of a physiological redox sensor that mediates leukocyte wound attraction in multicellular organisms.

Paracrine signalling by H₂O₂ induces neutrophil attraction to wounds⁶. However, what sensor detects H₂O₂ and mediates neutrophil recruitment is not known. H₂O₂ can cross cell membranes and inactivate tyrosine phosphatases through oxidation of the catalytic cysteine^{7–11}. Cysteine oxidation also regulates protein kinases^{12–15}. We used zebrafish, which are a powerful system to study vertebrate immunity^{6,16–19}, to identify a mechanism by which neutrophils detect wound-induced H₂O₂. While searching for a neutrophil redox sensor, we found that SFKs are activated in neutrophils around wounds. We detected autophosphorylation of the activation loop tyrosine of SFKs at 30 min after tail transection in 3 days post-fertilization (dpf) larvae (Fig. 1a–c). Phosphorylated SFKs showed a punctate appearance at the neutrophil leading edge and the autophosphorylation depended on wounding (Fig. 1b, c and Supplementary Fig. 2a, b). SFKs maintain inhibitory intramolecular interactions in an inactive state, while dephosphorylation of the carboxy-terminal phosphotyrosine releases the inhibitory configuration, allowing trans-autophosphorylation of the activation loop tyrosine, thereby activating SFKs^{20–23}. Additionally, emerging evidence suggests cysteine redox-mediated regulation of SFKs^{13,15}. Several cysteines are implicated in the redox regulation of cellular SRC (cSRC)^{13,15}, but redox regulation of other SFKs is poorly understood.

To investigate whether H₂O₂ is involved in SFK autophosphorylation in neutrophils, we inhibited Nox enzymes with diphenyleneiodonium (DPI)⁶. DPI attenuated SFK phosphorylation in neutrophils after wounding (Supplementary Fig. 2c, d). To distinguish H₂O₂ production from neutrophils versus wounds, we used morpholino antisense oligonucleotides to interfere with pre-mRNA splicing of dual oxidase (*duox*), which is responsible for H₂O₂ generation at wounds,

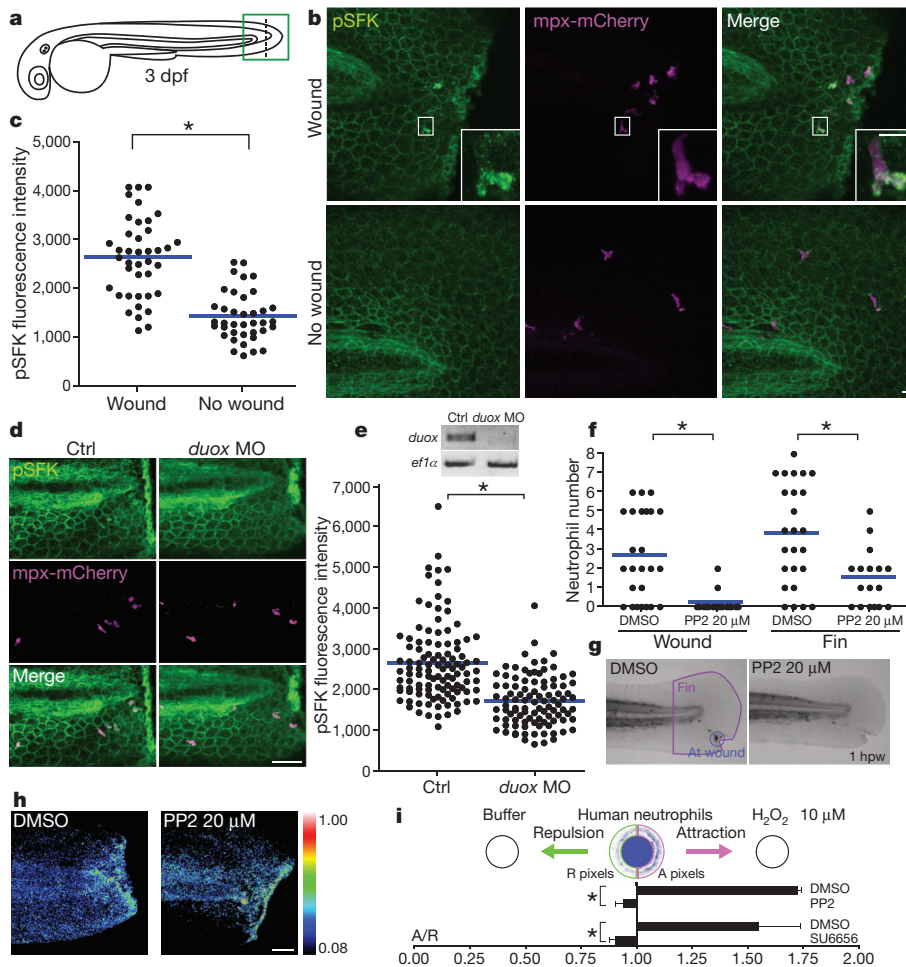
but not in neutrophils⁶. *Duox* knockdown inhibited SFK phosphorylation in neutrophils (Fig. 1d, e), indicating that SFK phosphorylation depends on the burst of H₂O₂ at wounds.

We found that treatment with SFK inhibitors impaired early accumulation of neutrophils at wounds (Fig. 1f, g and Supplementary Figs 2e, f and 3a, b). We focused on early recruitment of neutrophils to wounds at 0.5–1 h after wounding throughout this study, because within this timeframe the H₂O₂ burst occurs⁶ (Supplementary Movie 1) and neutrophil accumulation is roughly linear^{18,19}. Addition of the Src kinase inhibitor 4-amino-5-(4-chlorophenyl)-7-(dimethylethyl)pyrazolo[3,4-d]pyrimidine (PP2) at 1 h after wounding did not disturb resolution of inflammation, which is mediated by neutrophil reverse migration from wounds^{16,19} (Supplementary Fig. 3c, d). SFK inhibition did not impair the H₂O₂ burst at wounds (Fig. 1h and Supplementary Movie 1) or neutrophil motility in the cephalic mesenchyme¹⁸ (Supplementary Fig. 2g, h and Supplementary Movie 2). Injection of H₂O₂ into the otic cavity, or bath application of H₂O₂ after wounding, recruited neutrophils, and this was impaired by SFK inhibition (Supplementary Fig. 4a–f). We also performed an *in vitro* chemotaxis assay using human neutrophils. Consistent with mouse neutrophils⁷, H₂O₂ directly attracted human neutrophils (Fig. 1i). Two structurally different SFK inhibitors impaired chemotaxis to H₂O₂ (Fig. 1i) whereas SFK inhibition enhanced chemotaxis to N-formyl-methionyl-leucyl-phenylalanine (fMLP) (Supplementary Fig. 4g).

To identify specific SFKs that mediate neutrophil wound responses, we purified zebrafish myeloid cells by fluorescence-activated cell sorting (FACS) and performed reverse transcription–polymerase chain reaction (RT–PCR)¹⁹ (Supplementary Fig. 5a). Among the nine SFKs in mammals, LYN, FGR and HCK are myeloid-specific^{20,24}. Analysis of expressed sequence tag (EST) profiles in UniGene (NCBI) suggested that *lyn*, *hck*, *yrc* and *src* might be expressed in zebrafish leukocytes. RT–PCR detected *lyn* in neutrophils and *lyn* and *yrc* in macrophages (Fig. 2a). *In situ* hybridization detected *lyn* mRNA in the caudal haematopoietic tissue as well as in the pronephric ducts and neuromasts (Fig. 2b). LYN, which is phylogenetically conserved (Supplementary Fig. 5b), is expressed in all leukocytes except T lymphocytes in mammals, and its deficiency in mice leads to myeloproliferation and autoimmunity^{24–26}. *Lyn*-deficient neutrophils show an exaggerated response to adhesion signals and a weakened response to several chemokines *in vitro*, but the role of LYN in neutrophils *in vivo* is not clear^{24,27}. To knockdown zebrafish Lyn, we designed two morpholinos to target pre-mRNA splice sites (Fig. 2c). Both morpholinos attenuated neutrophil accumulation at wounds (Fig. 2d). Lyn knockdown increased the number of total neutrophils in the caudal haematopoietic tissue (Supplementary Fig. 5c, d), consistent with the myeloproliferation of *Lyn*-deficient mice^{24–26}. Live imaging demonstrated that Lyn depletion impaired neutrophil directional migration towards wounds (Fig. 2e–g and Supplementary Movie 3). Macrophage wound attraction was inhibited by PP2 or DPI but not Lyn knockdown (Supplementary Fig. 6a, b). Lyn knockdown also inhibited neutrophil responses to bath

¹Program in Cellular and Molecular Biology, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. ²Department of Medical Microbiology and Immunology, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA. ³Department of Pediatrics, University of Wisconsin-Madison, Madison, Wisconsin 53706, USA.

*These authors contributed equally to this work.



application of H_2O_2 (Supplementary Fig. 7a, b). To investigate specificity, we used leukotriene B₄ (LTB₄), which attracts zebrafish neutrophils¹⁷. Bath application of LTB₄ induced dissemination of neutrophils into fins, which was not prevented by either DPI or Lyn knockdown (Supplementary Fig. 7c–f and Supplementary Movie 4). Lyn depletion also inhibited the wound-induced SFK autophosphorylation in neutrophils (Fig. 2h, i), indicating that Lyn is primarily responsible for SFK phosphorylation in neutrophils.

We proposed that Lyn might be a neutrophil redox sensor that detects H_2O_2 at wounds. When we expressed Lyn fused to GFP at its C terminus in HEK293 cells, which produce H_2O_2 via NOX4 (ref. 28), phosphorylation of the Lyn activation loop tyrosine was detected (Fig. 3a). This autophosphorylation of Lyn partly depended on endogenous reactive oxygen species (ROS), because NOX inhibition by DPI decreased Lyn phosphorylation (Fig. 3a, b). Because endogenous ROS could also work on phosphatases and other kinases, we took a Lyn-specific approach. We created two cysteine mutants, C224A and C466A, mutating each cysteine to alanine on the basis of the importance of the corresponding residues in cSRC¹³ and phylogenetic conservation. When expressed in HEK293 cells, C466A had impaired autophosphorylation whereas C224A was phosphorylated similarly to Lyn wild type (Fig. 3c, d). C466 is located at the C terminus of the kinase domain and is phylogenetically conserved (Supplementary Fig. 8a). To investigate whether H_2O_2 activates Lyn directly through C466 oxidation, we performed an *in vitro* kinase assay. After immunoprecipitation of Lyn, we added 15 μM H_2O_2 , which is consistent with concentrations that attract neutrophils *in vitro*³ (Fig. 1i) and *in vivo*⁶. H_2O_2 directly activated Lyn wild type but not C466A (Fig. 3e, f). Importantly, C466A was activated by magnesium, which is necessary for the kinase reaction with ATP, similarly to wild type (Fig. 3g, h),

suggesting that C466 regulates the redox-specific function of Lyn rather than the basal activity. We also detected reduced cysteines in Lyn by biotinylated iodoacetamide (BIAM) labelling¹³ and found that the C466A mutant was slightly but significantly more reduced than wild type (Supplementary Fig. 8b, c), which is consistent with the idea that Lyn has multiple cysteines that could be oxidized. Because mitogen-activated protein kinases (MAPKs) are activated downstream of ROS and SFKs²⁹, we investigated the effects of Lyn expression on MAPKs. Ectopic expression of Lyn in HEK293 cells, in which Lyn is activated by endogenous ROS (Fig. 3a, b), activated ERK (also known as MAPK1) but not JNK (also known as MAPK8) or p38 (Supplementary Fig. 8d). This Lyn-induced activation of ERK was attenuated in cells expressing Lyn C466A compared to Lyn wild type (Supplementary Fig. 8e, f). Collectively, our data indicate that H_2O_2 directly activates Lyn through oxidation of C466, which activates downstream signalling. We also found that C468 in human LYN, which corresponds to C466 of zebrafish Lyn, has similar functions in regulating LYN autophosphorylation and oxidation and ERK activation (Supplementary Fig. 9a–e).

To investigate the importance of Lyn C466 during neutrophil wound responses, we established two transgenic zebrafish lines, which express Lyn (wild type)–GFP and Lyn(C466A)–GFP at similar levels in neutrophils (Supplementary Fig. 10a–d). We injected *lyn* morpholino, which does not inhibit the transgenes, into these transgenic animals. *lyn* morpholino did not inhibit neutrophil accumulation at wounds in transgenic larvae expressing Lyn wild type in neutrophils, although it impaired neutrophil recruitment in wild-type clutchmates (Fig. 4a, b), indicating that Lyn regulates neutrophil wound responses cell autonomously. When we injected *lyn* morpholino in the transgenic line expressing Lyn C466A in neutrophils, Lyn knockdown inhibited neutrophil wound responses in the transgenic larvae (Fig. 4c, d). These

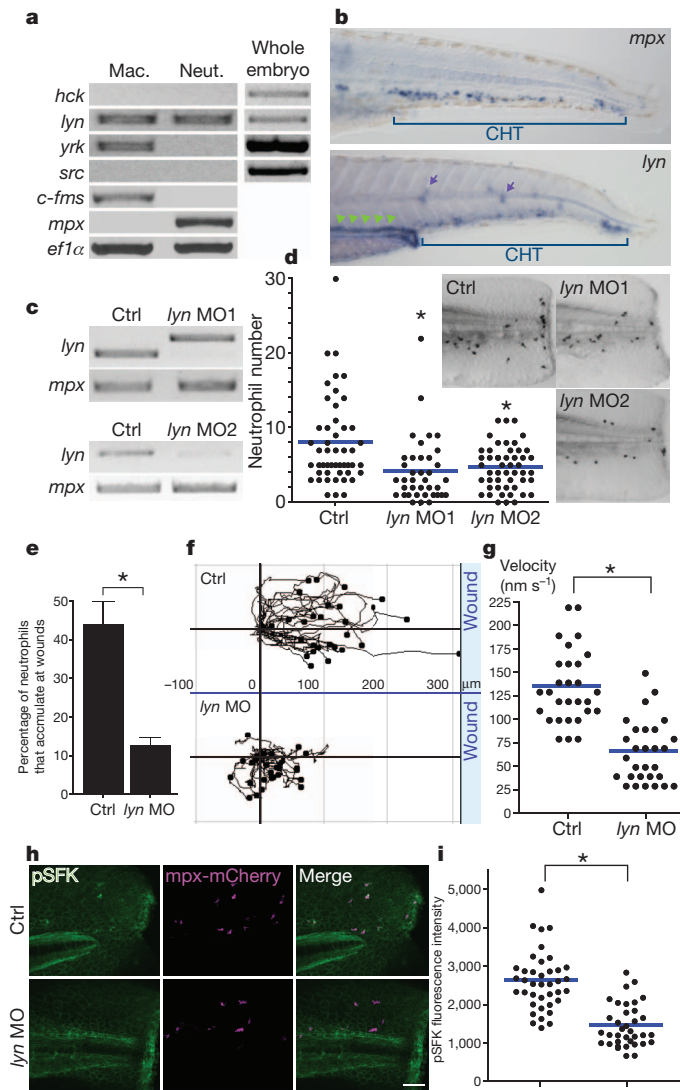


Figure 2 | Lyn mediates neutrophil wound responses. **a**, RT-PCR of SFKs. *c-fms* and *mpk* are markers of macrophages (Mac.) and neutrophils (Neut.), respectively, and *ef1α* is a loading control. **b**, *In situ* hybridization of *mpk* and *lyn* mRNA. Green arrowheads, pronephric ducts; purple arrows, neuromasts. CHT, caudal haematopoietic tissue. **c**, RT-PCR of *lyn* with or without *lyn* morpholinos. **d**, Neutrophil wound recruitment with or without *lyn* morpholinos at 30 min after wounding (control: 49 larvae; *lyn* MO1: 38 larvae; *lyn* MO2: 50 larvae). Pictures show Sudan Black staining. **e**, The percentage of neutrophils that accumulate at wounds 30 min after wounding (14 larvae each). **f**, Neutrophil tracking for 25 min after wounding (29 cells, 6 larvae each). **g**, Neutrophil velocity during wound responses (29 cells, 6 larvae each). **h**, Immunofluorescence of pSFK in *Tg(mpk:mCherry)* with or without *lyn* morpholino. **i**, Quantification of fluorescence intensity of pSFK in neutrophils (control: 39 cells (4 larvae); *lyn* MO: 36 cells (5 larvae)). Error bars indicate s.e.m. **P* < 0.05, one-way ANOVA with Dunnett post-test (**d**) and two-tailed unpaired *t*-test (**e**, **g**, **i**). Scale bar, 50 μm.

data demonstrate that C466, which is necessary for H_2O_2 -induced Lyn activation, mediates neutrophil responses *in vivo*. We also investigated the importance of the cysteine to signalling downstream of Lyn in neutrophils *in vivo*. Erk phosphorylation was detected in puncta at the leading edge of neutrophils around wounds (Supplementary Fig. 11a), which is reminiscent of phosphorylated SFKs (Fig. 1b). Lyn knockdown impaired Erk activation in neutrophils (Supplementary Fig. 11b, c), which was reversed by transgenic expression of Lyn wild type in neutrophils but not the C466A mutant, supporting the importance of C466 *in vivo* (Supplementary Fig. 11d–g). Furthermore,

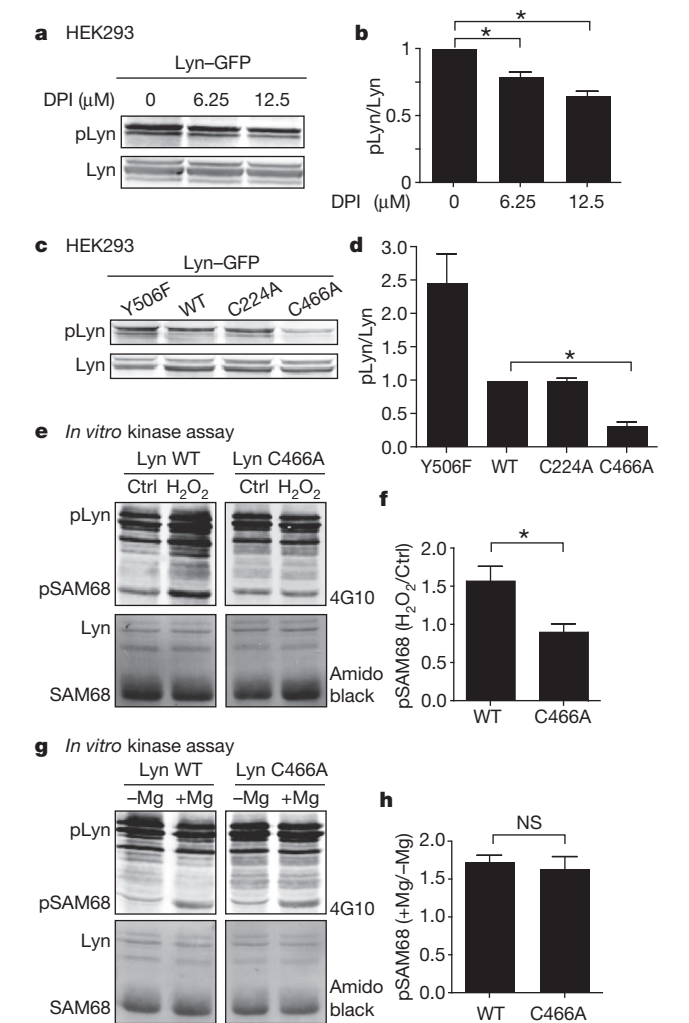


Figure 3 | H_2O_2 activates Lyn in a C466-dependent manner.

a, Autophosphorylation of Lyn activation loop tyrosine with or without DPI in HEK293 cells. **b**, Quantification of **a** (*n* = 4). **c**, Autophosphorylation of Lyn mutants in HEK293 cells. Y506F is a constitutively active mutant as a positive control for phospho-Lyn (pLyn). **d**, Quantification of **c** (*n* = 3). **e**, *In vitro* kinase assay of Lyn wild type (WT) and C466A with or without H_2O_2 using SAM68 as a substrate. **f**, Quantification of **e** (*n* = 6). **g**, *In vitro* kinase assay of Lyn wild type and C466A with or without magnesium. **h**, Quantification of **g** (*n* = 5). Error bars indicate s.e.m. **P* < 0.05, one-way ANOVA with Dunnett post-test (**b**), Bonferroni post-test (**d**) and two-tailed unpaired *t*-test (**f**, **h**). NS, not significant.

Erk inhibition impaired neutrophil wound attraction (Supplementary Fig. 12).

We found that Lyn acts as a direct redox sensor that mediates neutrophil wound responses *in vivo* (Supplementary Fig. 1a). Oxidation of cSRC regulates its activity, but whether it leads to inhibition or activation has been controversial^{13,15}. Oxidative stress activates LYN in neutrophils, but whether activation is direct or indirect was not known³⁰. Here we have shown that Lyn is activated through direct oxidation of C466 and elucidated the *in vivo* relevance of this. Because we did not detect oligomerization of Lyn in non-reducing SDS–polyacrylamide gel electrophoresis (SDS–PAGE; data not shown), oxidation at C466 might regulate an intramolecular conformation. H_2O_2 can cross the plasma membrane through aquaporins^{8,11}. How redox sensing by Lyn is transformed into directional migration remains to be determined. Activation of Lyn and Erk at the neutrophil front suggests that Lyn might regulate the neutrophil leading edge. LYN is unique in its dual roles in regulating both inhibitory signalling by phosphatases such as SHIP and activating signalling by MAPKs^{24–26} (Supplementary Fig. 1b).

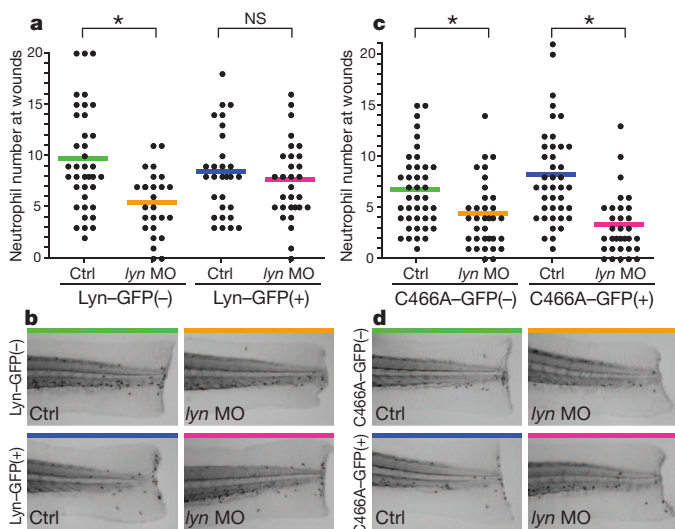


Figure 4 | Lyn regulates neutrophil wound responses in a C466-dependent manner. **a**, Neutrophil wound recruitment with or without *lyn* morpholino in *Tg(mpx:lyn-GFP)* and wild-type clutchmates (Lyn-GFP(-)/ctrl: 38 larvae; Lyn-GFP(-)/lyn MO: 29 larvae; Lyn-GFP(+)/ctrl: 26 larvae; Lyn-GFP(+)/lyn MO: 30 larvae). **b**, Representative pictures of Sudan Black staining in **a**. **c**, Neutrophil wound recruitment with or without *lyn* morpholino in *Tg(mpx:lyn C466A-GFP)* and wild-type clutchmates (Lyn(C466A)-GFP(-)/ctrl: 41 larvae; Lyn(C466A)-GFP(-)/lyn MO: 41 larvae; Lyn(C466A)-GFP(+)/ctrl: 34 larvae; Lyn(C466A)-GFP(+)/lyn MO: 32 larvae). C466A-GFP indicates Lyn(C466A)-GFP. **d**, Representative pictures of Sudan Black staining in **c**. * $P < 0.05$, one-way ANOVA with Bonferroni post-test (**a**, **c**).

How the opposing positive and negative signalling is coordinated is poorly understood. Because the phosphatases comprising LYN's inhibitory pathway are inactivated by oxidation^{9,10}, we speculate that H₂O₂ may have two effects on Lyn's function to orchestrate neutrophil wound responses: oxidation-mediated activation of Lyn and inhibition of the redox-sensitive inhibitory signalling downstream of Lyn (Supplementary Fig. 1b).

METHODS SUMMARY

Adult AB zebrafish and larvae were maintained as described previously¹⁸. For live imaging or wounding assays, larvae were anaesthetized in E3 containing 0.2 mg ml⁻¹ tricaine. To prevent pigment formation, some larvae were maintained in E3 containing 0.2 mM N-phenylthiourea. For morpholino experiments, 2.5–3 dpf larvae were used. For experiments without morpholinos, 3–3.5 dpf larvae were used. Confocal immunofluorescence images were acquired with a confocal microscope (FluoView FV1000, Olympus) using an NA 0.75/20× objective or an NA 1.45/60× oil immersion objective lens. Morpholino oligonucleotides (Gene Tools) in Danieau buffer were injected into 1-cell-stage embryos. *duox* splice morpholino (100 μM) was injected with *p53* morpholino (300 μM) as previously described⁶. *lyn* splice morpholino *lyn* MO1 (5'-TCAGACAGCAAATAGTAATCACCTT-3') (500 μM) or *lyn* splice morpholino *lyn* MO2 (5'-GAGTCTGTATTTCAG TACCATTAGC-3') (500 μM) was used (Danieau buffer as a control). To generate transgenic lines, 1 nl of solution containing 12.5 ng μl⁻¹ DNA plasmid (*tol2-mpx-lyn-GFP* or *tol2-mpx-lyn C466A-GFP*) and 17.5 ng μl⁻¹ transposase mRNA was injected into the cytoplasm of 1-cell-stage embryos. Injected embryos were raised to sexual maturity and screened by crossing with wild-type fish to identify founders. F1 embryos were identified by GFP expression and raised to sexual maturity. Experiments were performed on progeny of F1 or F2 outcrosses with wild-type fish. HEK293 cells were cultured in DME containing 10% FCS. Cells were transfected using Lipofectamine 2000 (Invitrogen).

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 20 May; accepted 14 October 2011.

Published online 20 November 2011.

1. Nathan, C. Neutrophils and immunity: challenges and opportunities. *Nature Rev. Immunol.* **6**, 173–182 (2006).

2. Dvorak, H. F. Tumors: wounds that do not heal. Similarities between tumor stroma generation and wound healing. *N. Engl. J. Med.* **315**, 1650–1659 (1986).
3. Klyubin, I. V., Kirpichnikova, K. M. & Gamaley, I. A. Hydrogen peroxide-induced chemotaxis of mouse peritoneal neutrophils. *Eur. J. Cell Biol.* **70**, 347–351 (1996).
4. Feng, Y., Santoriello, C., Mione, M., Hurlstone, A. & Martin, P. Live imaging of innate immune cell sensing of transformed cells in zebrafish larvae: parallels between tumor initiation and wound inflammation. *PLoS Biol.* **8**, e1000562 (2010).
5. Moreira, S., Stramer, B., Evans, I., Wood, W. & Martin, P. Prioritization of competing damage and developmental signals by migrating macrophages in the *Drosophila* embryo. *Curr. Biol.* **20**, 464–470 (2010).
6. Niethammer, P., Grabher, C., Look, A. T. & Mitchison, T. J. A tissue-scale gradient of hydrogen peroxide mediates rapid wound detection in zebrafish. *Nature* **459**, 996–999 (2009).
7. Rhee, S. G. Cell signaling. H₂O₂, a necessary evil for cell signaling. *Science* **312**, 1882–1883 (2006).
8. Bienert, G. P., Schjoerring, J. K. & Jahn, T. P. Membrane transport of hydrogen peroxide. *Biochim. Biophys. Acta* **1758**, 994–1003 (2006).
9. Paulsen, C. E. & Carroll, K. S. Orchestrating redox signaling networks through regulatory cysteine switches. *ACS Chem. Biol.* **5**, 47–62 (2010).
10. Poole, L. B. & Nelson, K. J. Discovering mechanisms of signaling-mediated cysteine oxidation. *Curr. Opin. Chem. Biol.* **12**, 18–24 (2008).
11. Miller, E. W., Dickinson, B. C. & Chang, C. J. Aquaporin-3 mediates hydrogen peroxide uptake to regulate downstream intracellular signaling. *Proc. Natl Acad. Sci. USA* **107**, 15681–15686 (2010).
12. Burgoyne, J. R. *et al.* Cysteine redox sensor in PKGα enables oxidant-induced activation. *Science* **317**, 1393–1397 (2007).
13. Giannoni, E., Buricchi, F., Rauegi, G., Ramponi, G. & Chiarugi, P. Intracellular reactive oxygen species activate Src tyrosine kinase during cell adhesion and anchorage-dependent cell growth. *Mol. Cell Biol.* **25**, 6391–6403 (2005).
14. Guo, Z., Kozlov, S., Lavin, M. F., Person, M. D. & Paul, T. T. ATM activation by oxidative stress. *Science* **330**, 517–521 (2010).
15. Kemble, D. J. & Sun, G. Direct and specific inactivation of protein tyrosine kinases in the Src and FGFR families by reversible cysteine oxidation. *Proc. Natl Acad. Sci. USA* **106**, 5070–5075 (2009).
16. Mathias, J. R. *et al.* Resolution of inflammation by retrograde chemotaxis of neutrophils in transgenic zebrafish. *J. Leukoc. Biol.* **80**, 1281–1288 (2006).
17. Tobin, D. M. *et al.* The *Ita4h* locus modulates susceptibility to mycobacterial infection in zebrafish and humans. *Cell* **140**, 717–730 (2010).
18. Yoo, S. K. *et al.* Differential regulation of protrusion and polarity by PI3K during neutrophil motility in live zebrafish. *Dev. Cell* **18**, 226–236 (2010).
19. Yoo, S. K. & Huttenlocher, A. Spatiotemporal photolabeling of neutrophil trafficking during inflammation in live zebrafish. *J. Leukoc. Biol.* **89**, 661–667 (2011).
20. Martin, G. S. The hunting of the Src. *Nature Rev. Mol. Cell Biol.* **2**, 467–475 (2001).
21. Sicheri, F., Moarefi, I. & Kuriyan, J. Crystal structure of the Src family tyrosine kinase Hck. *Nature* **385**, 602–609 (1997).
22. Xu, W., Harrison, S. C. & Eck, M. J. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* **385**, 595–602 (1997).
23. Yeatman, T. J. A renaissance for SRC. *Nature Rev. Cancer* **4**, 470–480 (2004).
24. Scapini, P., Pereira, S., Zhang, H. & Lowell, C. A. Multiple roles of Lyn kinase in myeloid cell signaling and function. *Immunol. Rev.* **228**, 23–40 (2009).
25. Hibbs, M. L. *et al.* Multiple defects in the immune system of Lyn-deficient mice, culminating in autoimmune disease. *Cell* **83**, 301–311 (1995).
26. Nishizumi, H. *et al.* Impaired proliferation of peripheral B cells and indication of autoimmune disease in Lyn-deficient mice. *Immunity* **3**, 549–560 (1995).
27. Pereira, S. & Lowell, C. The Lyn tyrosine kinase negatively regulates neutrophil integrin signaling. *J. Immunol.* **171**, 1319–1327 (2003).
28. Lee, Y. M. *et al.* NOX4 as an oxygen sensor to regulate TASK-1 activity. *Cell. Signal.* **18**, 499–507 (2006).
29. Abe, J., Takahashi, M., Ishida, M., Lee, J. D. & Berk, B. C. c-Src is required for oxidative stress-mediated activation of big mitogen-activated protein kinase 1. *J. Biol. Chem.* **272**, 20389–20394 (1997).
30. Yan, S. R. & Berton, G. Regulation of Src family tyrosine kinase activities in adherent human neutrophils. Evidence that reactive oxygen intermediates produced by adherent neutrophils increase the activity of the p58^{c-fgr} and p53/56^{lyn} tyrosine kinases. *J. Biol. Chem.* **271**, 23464–23471 (1996).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank K. T. Chan for help with tissue culture work, J. M. Green and P.-Y. Lam for help with *in situ* hybridization, M. Shelef and S. Wernimont for drawing blood, and A. J. Wiemer for insightful discussion and critical reading of the manuscript. This work was supported by American Heart Association fellowship 11PRE4890041 (S.K.Y.), National Institutes of Health Grant GM074827 (A.H.), NIH Research Training Grant in Hematology 5T32 HL07899 (T.W.S.) and UW MSTP (T.W.S.).

Author Contributions S.K.Y. designed the research, performed the experiments, analysed data and wrote the paper. T.W.S. contributed to development and data analysis of the *in vitro* chemotaxis assay. Q.D. constructed the HyPer probe and contributed expertise in zebrafish injection. A.H. designed the research, analysed data and co-wrote the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to A.H. (huttenlocher@wisc.edu).

METHODS

Zebrafish maintenance and general procedures. Adult AB zebrafish and larvae were maintained as described previously¹⁸. For live imaging or wounding assay, larvae were anaesthetized in E3 containing 0.2 mg ml⁻¹ tricaine (ethyl 3-aminobenzoate, Sigma Aldrich). To prevent pigment formation, some larvae were maintained in E3 containing 0.2 mM N-phenylthiourea (PTU, Sigma Aldrich). For morpholino experiments, 2.5–3 dpf larvae were used. For experiments without morpholinos, 3–3.5 dpf larvae were used. When drugs were used, larvae were pre-treated at least for 1 h before experiments unless otherwise indicated (10–20 µM PP2, 10 µM PP1, 50–100 µM DPI, 100 µM PD98059, 20 µM U0126). Tail transection was performed on 2.5–3 dpf larvae using a razor blade. Tail fins of 3–3.5 dpf larvae were wounded with a needle. For application of exogenous H₂O₂, we injected one nanolitre of 10 µM H₂O₂ into the otic cavity of 3 dpf larvae or bathed 2.5–3 dpf larvae in E3 containing 100 µM H₂O₂ after tail transection. For bath application of LTB₄, 2.5–3 dpf larvae were bathed in E3 containing 30 µM LTB₄.

Immunofluorescence and Sudan Black staining. Thirty minutes after tail transection, 2.5–3 dpf larvae were fixed with 1.5% formaldehyde in 0.1 M PIPES, 1.0 mM MgSO₄ and 2 mM EGTA overnight at 4 °C and immunolabelled as previously described¹⁹. We used the following primary antibodies: rabbit anti-Mpx antibody^{19,31} at 1:300, rabbit anti-L-Plastin antibody^{19,31} at 1:300, rabbit anti-phospho-Src family (Tyr416) antibody (Cell Signaling, #2101) at 1:300, rabbit anti-phospho-Src family (Tyr416) antibody (Cell Signaling, D49G4) at 1:300, mouse anti-phospho-Erk1/2 (T185, Y187, T202, Y204) antibody (Abcam, ab50011) at 1:300. For pSFKs staining, the rabbit anti-phospho-Src family (Tyr416) antibody (Cell Signaling, #2101) was used except that presented in Supplementary Fig. 2a, which was stained with the rabbit anti-phospho-Src family (Tyr416) antibody (Cell Signaling, D49G4). DyLight 488- or 549-conjugated IgG antibodies (Jackson ImmunoResearch) were used as secondary antibodies. Sudan Black staining was performed as previously described^{18,19}.

FACS and RT-PCR. Dissociated cells of *Tg(mpx:Dendra2)* at 3 dpf were sorted by FACS as previously described¹⁹. RNA was isolated using the RNeasy Mini Kit (Qiagen) and one-step RT-PCR (Qiagen) was performed. Primers for *c-fms*, *mpx* and *efl1x* have been previously described³¹. Other oligonucleotide sequences were as follows: *hck* forward, 5'-TGCCAGTCTCTACCTCTCC-3'; *hck* reverse, 5'-GCCGAAAGACCACACATCGG-3'; *lyn* forward, 5'-CGAAAGCTGGAT AAAGCATGCG-3'; *lyn* reverse, 5'-CTGCTCTCAGGTCTCGGTGG-3'; *yrk* forward, 5'-CAGATCATGAAGAGGCTCCGTC-3'; *yrk* reverse, 5'-CCTGCTC CAGCAGCTCTCGG-3'; *src* forward, 5'-ACGGCTGTGCCACTCCCTG-3'; *src* reverse, 5'-CCGTACAGAGCTGCCTCCGG-3'.

Morpholino injection and RT-PCR. Morpholino oligonucleotides (Gene Tools) in Danieau buffer (58 mM NaCl, 0.7 mM KCl, 0.4 mM MgSO₄, 0.6 mM Ca(NO₃)₂, 5.0 mM HEPES pH 7.1–7.3) were injected (1 nl) into 1-cell-stage embryos. *duox* splice morpholino (5'-AGTGAATTAGAGAAATGCACCTTTT-3') (100 µM) was injected with *p53* morpholino (5'-GCGCCATTGCTTGTCAAGAATTG-3') (300 µM) as previously described (*p53* morpholino (300 µM) was used as a control)⁶. *lyn* splice morpholino *lyn* MO1 (5'-TCAGACAGCAAATAGTAATCA CCTT-3') (500 µM) or *lyn* splice morpholino *lyn* MO2 (5'-GAGTCTGTAT TTCATACCATAGC-3') (500 µM) was used (Danieau buffer as a control). For morphotyping of the splicing morpholinos, RNA was prepared from 2.5–3 dpf larvae using TRIzol (Invitrogen) and one-step RT-PCR (Qiagen) was performed. Primers for *duox* have been described previously⁶. Other oligonucleotide sequences were as follows: *lyn* forward (for *lyn* MO1), 5'-GCGAGGTCTTT GACCACACG-3'; *lyn* reverse (for *lyn* MO1), 5'-CGCATGCTTTATCCAGC TTTCGAC-3'; *lyn* forward (for *lyn* MO2), 5'-CGAAAGCTGGATATAAAGC ATGCG-3'; *lyn* reverse (for *lyn* MO2), 5'-CTGCTCTCAGGTCTCGGTGG-3'.

Whole-mount *in situ* hybridization. For *in situ* hybridization, 1 kb of the 3' region of zebrafish *lyn* was amplified with the T7 promoter by RT-PCR (5'-CGTAGACGTCAGGGCCAGG-3' and 5'-GATCAATAACGACTCAC TATAGGTCTAGGCTGCTGCTGGTACTGG-3') using mRNA from 3 dpf larvae as a template. Digoxigenin-labelled RNA probe was transcribed with the use of T7 RNA polymerase (Ambion). The *mpx* probe was transcribed as previously described³². 3 dpf larvae were fixed in 4% paraformaldehyde in PBS and mRNA was labelled by *in situ* hybridization using the previously described protocol³³.

Image acquisition and analysis. Confocal immunofluorescence images were acquired with a confocal microscope (FluoView FV1000, Olympus) using an NA 0.75/20× objective or an NA 1.45/60× oil immersion objective lens. Z-series were acquired using a 260–600-µm pinhole. For quantification of fluorescence intensity of pSFK and pErk in neutrophils, the median intensity of a square with three-pixel sides was measured at a single z-plane and the highest median intensity within each neutrophil was used. For H₂O₂ imaging, HyPer fluorescence was excited with 405 nm and 488 nm lasers, and emission wavelengths 505–510 and 510–525 nm (dichroic mirror: SDM510; band pass filter: BA505–525) were

acquired using the sequential line scanning. Ratiometric analysis was performed by using FluoView FV1000 software (Olympus) after Z-series stacking as previously described¹⁸. Neutrophils were tracked and analysed by using plugins MTrackj (three-dimensional tracking), Manual tracking (two-dimensional tracking) and Chemotaxis and Migration tool (Ibidi) for ImageJ (National Institutes of Health). The percentage of neutrophils that accumulate at wounds was calculated by dividing the number of neutrophils that arrived at a wound 30 min after wounding by the total number of neutrophils in a region that is 400–500 µm distant from the wound margin. A Nikon SMZ-1500 stereomicroscope (Nikon) was used for imaging of larvae labelled with Sudan Black staining, *in situ* hybridization and time-lapse analysis of *Tg(mpx:Dendra2)*.

Photolabelling of neutrophils. Neutrophils at wounds were photolabelled as previously described¹⁹. *Tg(mpx:Dendra2)* at 3 dpf were wounded with a razor blade. At 1 h after wounding, larvae were treated with DMSO or PP2 and neutrophils at wounds were photolabelled by photoconverting fluorescence of Dendra2 from green (488 nm) to red (543 nm). For photoconversion, a 405 nm laser was focused into an area of interest for 20–30 s with 20–30% power 10.0 µs pixel⁻¹ (tornado function). Photolabelled neutrophils were tracked up to 9.5 h after wounding.

Plasmid construction and *in vitro* transcription. The plasmid for zebrafish *Lyn* wild-type expression in HEK293 cells was constructed by subcloning codon-optimized *lyn* wild type (Supplementary Fig. 13a) into pEGFP-N1 vector (clontech). Humanization and optimization of codon usage was performed (GenScript) owing to poor expression of the original zebrafish *lyn* in HEK293 cells. *lyn* mutants (C466A, C224A and Y506F) in pEGFP-N1 were made by overlap PCR mutagenesis. Codon-optimized *lyn*-GFP wild type or *lyn*(C466A)-GFP was subcloned into the backbone vector with the *mpx* promoter¹⁶, minimal *tol2* elements³⁴ and a SV40 polyadenylation sequence as previously described^{18,19}. HyPer³⁵ was subcloned into the pCS2+ vector, linearized by NotI and *in vitro* transcribed by SP6 RNA polymerase (Invitrogen). The plasmid for human *LYN* wild-type expression in HEK293 cells was constructed by subcloning codon-optimized human *LYN* wild type (Supplementary Fig. 13b) into pEGFP-N1 vector (clontech). Human *LYN* C468A in pEGFP-N1 was made by overlap PCR mutagenesis.

Cells and transfection. HEK293 cells were cultured in DME containing 10% FCS. Cells were transfected using Lipofectamine 2000 (Invitrogen) according to the manufacturer's instruction. Cells were treated with DPI for 6–7 h before lysis (see Fig. 3a).

Immunoblot analysis. HEK293 cells were lysed in lysis buffer (50 mM HEPES, pH 7.4, 150 mM NaCl, 0.5% Nonidet P-40, 5% glycerol, 1 mM MgCl₂, 1 mM MnCl₂, 20 mM NaF, 1 mM Na₃VO₄, 1 mM dithiothreitol, 0.2 mM PMSF, 1 µg ml⁻¹ pepstatin, 10 µg ml⁻¹ aprotinin, 5 µg ml⁻¹ leupeptin, Phosphatase Inhibitor Cocktail 2 (Sigma) at 1:1,000). Whole-cell lysates were subjected to analyses by immunoblotting with an infrared imaging system (Odyssey; LI-COR Biosciences) as previously described³⁶. Antibodies against phospho-Src family (Tyr416) (Cell Signaling, #2101), GFP (Clontech, JL-8), phospho-ERK1/2 (Abcam, ab50011), ERK1/2 (BioSource, 44654G), phospho-ERK5 (Sigma, E7153), ERK5 (Santa Cruz, 12F2), phospho-JNK (Santa Cruz, G-7), JNK (BioSource, 44690), phospho-p38 (Calbiochem, 506124) and p38 (BioSource, 2F11) were purchased commercially.

Immunoprecipitation and *in vitro* kinase assay. Twenty-four hours after transfection of *Lyn*-GFP wild type or *Lyn*(C466A)-GFP, HEK293 cells were treated with 5 µM PP2 for 2–4 h to inactivate the basal activity of *Lyn*. Cells were lysed in cRIPA lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Triton X-100, 2 mM EGTA, 1 mM Na₃VO₄, 1 mM dithiothreitol, 0.2 mM PMSF, 10 µg ml⁻¹ aprotinin, 5 µg ml⁻¹ leupeptin, 1 µM PP2, 0.1% SDS, 0.5% sodium deoxycholate). *Lyn*-GFP was immunoprecipitated with anti-GFP serum (Invitrogen) and GammaBind G-Sepharose beads (GE Healthcare). The immune complex was washed twice with cRIPA buffer (5 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Triton X-100, 2 mM EGTA, 0.1% SDS, 0.5% sodium deoxycholate) and twice with kinase buffer (10 mM Tris-HCl, pH 7.5, 5 mM MgCl₂). The immunoprecipitate was treated with or without 15 µM H₂O₂ in kinase buffer (40 µl) containing 100 µM ATP and 0.5 µg SAM68 (ref. 37) (Santa Cruz) at room temperature (24 °C) for 45 s with continuous mixing³⁸. The reaction was stopped by addition of 5× sample buffer and the immunocomplexes were separated by SDS-PAGE and probed with anti-phosphotyrosine antibody (4G10 Platinum, Millipore). The precipitated *Lyn* and SAM68 were detected by Amido black staining. When we performed the kinase assay to examine effects of magnesium on *Lyn* activity (Fig. 3g, h), MgCl₂ was removed from kinase buffer for washing and the immunoprecipitate was treated with or without 5 mM MgCl₂ during the kinase assay.

BIAM-mediated carboxymethylation. Twenty-four hours after transfection, HEK293 cells were lysed in cRIPA lysis buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1% Triton X-100, 2 mM EGTA, 1 mM Na₃VO₄, 0.2 mM PMSF, 10 µg ml⁻¹ aprotinin, 5 µg ml⁻¹ leupeptin, 0.1% SDS, 0.5% sodium deoxycholate)

containing 100 μM BIAM (Invitrogen). Lyn-GFP was immunoprecipitated with anti-GFP serum (Invitrogen) and GammaBind G-Sepharose beads (GE Healthcare). The immunocomplexes were separated by SDS-PAGE and probed with anti-GFP antibody (JL-8, Clontech) and IRDye 680 Streptavidin (LI-COR Biosciences).

Generation of *Tg(mpx:lyn-GFP)* and *Tg(mpx:lyn C466A-GFP)*. One nanolitre of solution containing 12.5 $\text{ng } \mu\text{l}^{-1}$ DNA plasmid (*tol2-mpx-lyn-GFP* or *tol2-mpx-lyn C466A-GFP*) and 17.5 $\text{ng } \mu\text{l}^{-1}$ transposase mRNA was injected into the cytoplasm of 1-cell-stage embryos. Injected embryos were raised to sexual maturity and screened by crossing with wild-type fish to identify founders. F1 embryos were identified by GFP expression and raised to sexual maturity. Experiments were performed on progeny of F1 or F2 outcross with wild-type fish.

Purification of human neutrophils and *in vitro* chemotaxis assay. Blood was obtained from healthy donors with informed consent. Neutrophils were purified using Polymorphprep (Nycomed Pharma AS). Purified neutrophils were suspended in HBSS containing 5% FBS and 25 mM HEPES at $1.5\text{--}2.5 \times 10^7$ cells ml^{-1} . Cells were pretreated with DMSO, 10 μM PP2 or 10 μM SU6656 for 30 min. The under-agarose assay was performed as previously described³⁹ with minor modification. Plastic tissue culture dishes (35-mm diameter) were filled with 3 ml of a 1.2% agarose solution containing 50% HBSS and 50% RPMI-1640 with 20% FBS. After the agarose solidified, four series of three wells, 2 mm in diameter and spaced 2 mm apart, were made. The centre well of each three-well series was loaded with 20 μl of neutrophils ($1.5\text{--}2.5 \times 10^7$ cells ml^{-1}) and the outer wells were loaded with chemoattractants (10 μM H_2O_2 or 100 nM fMLP) or HBSS. Once loaded, the gels were incubated for 1 h in a 37 °C/5% CO_2 incubator. Cells were fixed with methanol for a few hours, and with 37% formaldehyde overnight. After fixation, the gels were carefully removed and cells were stained using a Hema-3 stain kit (Thermo Fisher Scientific). For quantification, four dishes were used for each condition. Cells were imaged using a Nikon SMZ-1500 stereomicroscope (Nikon) and acquired pictures were thresholded to generate binary images. Pixel intensities of binary images on each side of the centre well were quantified by MetaView software. The pixel intensity on the chemoattraction side was divided by the intensity on the buffer side. The geometric mean of 16 values (each of four dishes

has four series of three wells) was calculated for each condition. The means of three separate experiments were calculated and shown with s.e.m.

Statistics. All error bars indicate s.e.m. For quantification of western blot and *in vitro* chemotaxis assay, data shown are means \pm s.e.m. of at least three separate experiments. For quantification of animal experiments *in vivo*, representative data of two or three separate experiments are presented unless otherwise indicated. Assuming Gaussian distribution of overall population of values, *P* values were derived by the following analyses.

Two-tailed unpaired *t*-test: Figs 1c, e, f, i, 2e, g, i and 3f, h and Supplementary Figs 2b, d, e, h, 4g, 7c, e and 11c, f, g. Two-tailed paired *t*-test: Supplementary Figs 8c, f and 9c, e. One-way ANOVA with Bonferroni post-test: Figs 3d and 4a, c and Supplementary Figs 4c, f and 7a. One-way ANOVA with Dunnett post-test: Figs 2d and 3b and Supplementary Figs 2f, 5c, 6a and 12a.

31. Mathias, J. R. *et al.* Characterization of zebrafish larval inflammatory macrophages. *Dev. Comp. Immunol.* **33**, 1212–1217 (2009).
32. Bennett, C. M. *et al.* Myelopoiesis in the zebrafish, *Danio rerio*. *Blood* **98**, 643–651 (2001).
33. Thisse, C. & Thisse, B. High-resolution *in situ* hybridization to whole-mount zebrafish embryos. *Nature Protocols* **3**, 59–69 (2008).
34. Urasaki, A., Morvan, G. & Kawakami, K. Functional dissection of the *Tol2* transposable element identified the minimal *cis*-sequence and a highly repetitive sequence in the subterminal region essential for transposition. *Genetics* **174**, 639–649 (2006).
35. Belousov, V. V. *et al.* Genetically encoded fluorescent indicator for intracellular hydrogen peroxide. *Nature Methods* **3**, 281–286 (2006).
36. Chan, K. T., Cortesio, C. L. & Huttenlocher, A. FAK alters invadopodia and focal adhesion composition and dynamics to regulate breast cancer invasion. *J. Cell Biol.* **185**, 357–370 (2009).
37. Wong, B. R. *et al.* TRANCE, a TNF family member, activates Akt/PKB through a signaling complex involving TRAF6 and c-Src. *Mol. Cell* **4**, 1041–1049 (1999).
38. Yamanashi, Y. *et al.* Activation of Src-like protein-tyrosine kinase Lyn and its association with phosphatidylinositol 3-kinase upon B-cell antigen receptor-mediated signaling. *Proc. Natl Acad. Sci. USA* **89**, 1118–1122 (1992).
39. Heit, B. & Kubes, P. Measuring chemotaxis and chemokinesis: the under-agarose cell migration assay. *Sci. STKE* **2003**, pl5 (2003).

Image-based genome-wide siRNA screen identifies selective autophagy factors

Anthony Orvedahl^{1,2*}, Rhea Sumpter Jr^{1,3*}, Guanghua Xiao⁴, Aylwin Ng^{5,6,7}, Zhongju Zou^{1,3,8}, Yi Tang⁹, Masahiro Narimatsu¹⁰, Christopher Gilpin¹¹, Qihua Sun^{1,3}, Michael Roth^{12,13}, Christian V. Forst⁴, Jeffrey L. Wrana^{10,14}, Ying E. Zhang⁹, Katherine Luby-Phelps¹¹, Ramnik J. Xavier^{5,6,7}, Yang Xie^{4,13} & Beth Levine^{1,2,3,8,13}

Selective autophagy involves the recognition and targeting of specific cargo, such as damaged organelles, misfolded proteins, or invading pathogens for lysosomal destruction^{1–4}. Yeast genetic screens have identified proteins required for different forms of selective autophagy, including cytoplasm-to-vacuole targeting, pexophagy and mitophagy, and mammalian genetic screens have identified proteins required for autophagy regulation⁵. However, there have been no systematic approaches to identify molecular determinants of selective autophagy in mammalian cells. Here, to identify mammalian genes required for selective autophagy, we performed a high-content, image-based, genome-wide small interfering RNA screen to detect genes required for the colocalization of Sindbis virus capsid protein with autophagolysosomes. We identified 141 candidate genes required for viral autophagy, which were enriched for cellular pathways related to messenger RNA processing, interferon signalling, vesicle trafficking, cytoskeletal motor function and metabolism. Ninety-six of these genes were also required for Parkin-mediated mitophagy, indicating that common molecular determinants may be involved in autophagic targeting of viral nucleocapsids and autophagic targeting of damaged mitochondria. Murine embryonic fibroblasts lacking one of these gene products, the C2-domain containing protein, SMURF1, are deficient in the autophagosomal targeting of Sindbis and herpes simplex viruses and in the clearance of damaged mitochondria. Moreover, SMURF1-deficient mice accumulate damaged mitochondria in the heart, brain and liver. Thus, our study identifies candidate determinants of selective autophagy, and defines SMURF1 as a newly recognized mediator of both viral autophagy and mitophagy.

To identify novel genes required for selective autophagy, we performed a genome-wide siRNA screen to detect changes in the colocalization of a red-labelled Sindbis virus (SIN) capsid protein with a green fluorescent protein (GFP)-labelled marker of autophagosomes, GFP-LC3 (LC3 is also known as MAP1LC3) (Supplementary Fig. 1a) in SIN-infected HeLa/GFP-LC3 cells (ref. 6). Using correlative light and electron microscopy, we confirmed that colocalized red and green puncta represented autophagic structures (primarily autolysosomes) containing numerous viral nucleocapsids (Fig. 1a). The predominance of viral nucleocapsids concentrated in these structures (relative to within the cytoplasm) is consistent with selective autophagic targeting of viral nucleocapsids (herein referred to as virophagy).

Screening of a human siGenome library containing 21,125 siRNA pools showed that knockdown of 195 and 13 genes resulted in decreased

or increased colocalization, respectively, (Fig. 1b, Supplementary Table 1 and Supplementary Fig. 1b). Genes were re-screened with sets of four individual siRNAs (Supplementary Table 2; see column 'J' of Supplementary Table 3 for siRNA sequences) to confirm our primary screen and rule out potential off-target effects of individual siRNAs; knockdown with two or more siRNAs resulted in decreased colocalization for 141 (72%) genes. (Fig. 2 and Supplementary Figs 1b and 2a). None of these 141 gene knockdowns decreased numbers of green puncta in uninfected cells (data not shown), indicating that these genes function in virophagy, but not in regulation of autophagy. There was no enrichment of siRNAs-containing microRNA seed sequences among these genes ($P = 0.95$) (Supplementary Tables 3 and 4), indicating that bias due to miRNA-like off-target effects was unlikely. There was a low confirmation rate for siRNAs that increased colocalization (2 of 13 genes); therefore, we subsequently focused only on siRNAs that decreased colocalization.

Bioinformatic analyses of the 141 confirmed hits required for SIN capsid/GFP-LC3 colocalization showed enrichment for gene sets associated with biological processes and molecular functions including RNA splicing/processing, protein phosphorylation, transport, calcium-binding and the cytoskeleton (Supplementary Table 5 and Supplementary Fig. 3a). Examination of our hits within a framework of functional cellular pathways revealed strongly enriched network modules associated with RNA processing, interferon (IFN)- α and - γ signalling, SNARE vesicular transport, cytoskeletal-associated components, and several metabolic pathways (Supplementary Fig. 3b and Fig. 1c). This is consistent with the function of IFN- γ in selective microbial autophagy¹ and the described role of the actin cytoskeleton in selective autophagy in yeast⁷ and mammalian⁸ cells. The enrichment of SNARE proteins suggests that in addition to a function in autophagosome formation and maturation^{9,10}, these proteins may be involved in the trafficking of selective cargo to the autophagosome. Twelve colocalization hits form primary interactions with core autophagy machinery and associated components¹¹ (Fig. 1d). One colocalization screen hit, clathrin interaction 1 (CLINT) interacts with ATG8 components (GABARAPL1, MAP1LC3A, MAP1LC3B), which are crucial in the recognition of cargo during selective autophagy². Another hit, ATG13, is a member of the core autophagy network, indicating that it may have an as yet undefined function in selective autophagy, in addition to its role in the ULK1 (also known as Atg1) autophagy induction complex¹². Five colocalization hits, SMURF1, NEFM, KCNAB2, SFRS4 and UBA52, interact with p62 (also known as SQSTM1), a known adaptor in diverse

¹Department of Internal Medicine, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9113, USA. ²Department of Microbiology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9113, USA. ³Center for Autophagy Research, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9113, USA. ⁴Department of Clinical Sciences, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9113, USA. ⁵Center for Computational and Integrative Biology, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. ⁶Gastrointestinal Unit, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts 02114, USA. ⁷Broad Institute of Harvard and Massachusetts Institute of Technology, Cambridge, Massachusetts 02142, USA. ⁸Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9113, USA. ⁹Laboratory of Cellular and Molecular Biology, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, USA. ¹⁰Center for Systems Biology, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada. ¹¹Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9113, USA. ¹²Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9113, USA. ¹³Harold C. Simmons Comprehensive Cancer Center, University of Texas Southwestern Medical Center, Dallas, Texas 75390-9113, USA. ¹⁴Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 3E1, Canada.

*These authors contributed equally to this work.

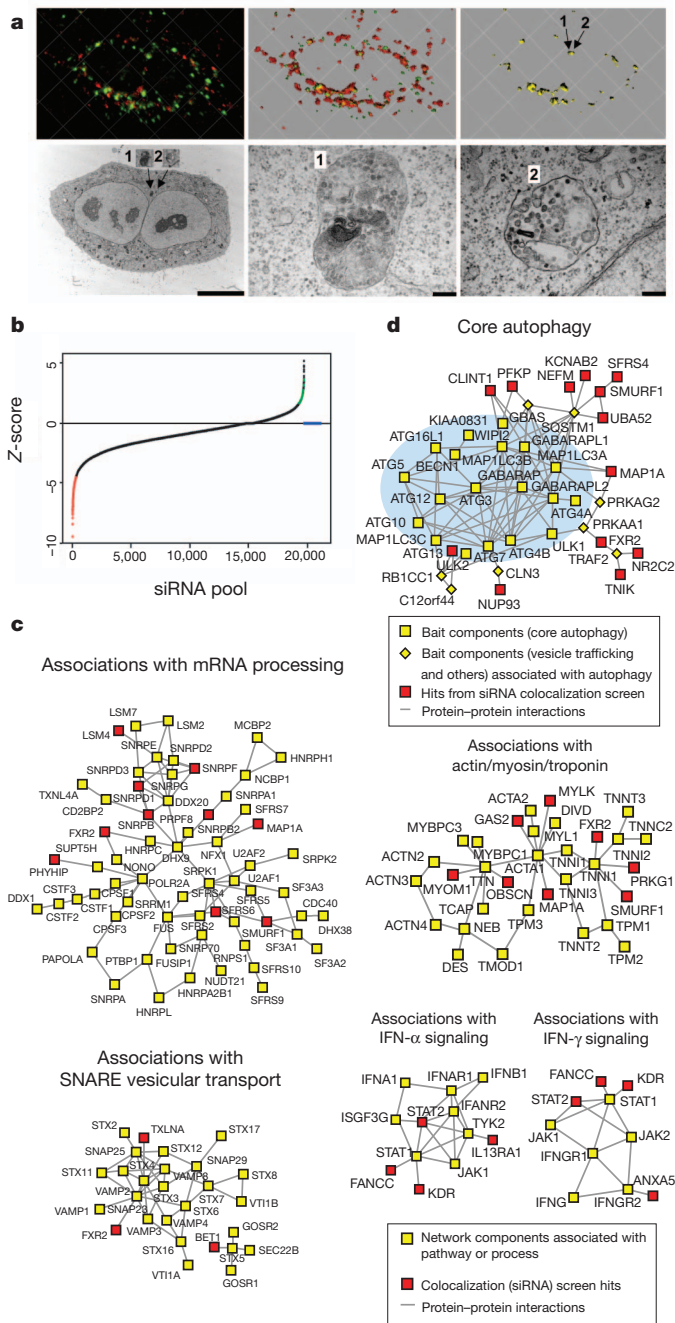


Figure 1 | Genome-wide screen to identify cellular factors required for Sindbis virus capsid colocalization with autophagosomes. **a**, Correlative light and electron microscopy of HeLa/GFP-LC3 cell infected with SIN/mCherry.capsid virus. Top left, deconvolved image of red and green fluorescence channels; middle, three-dimensional surface reconstruction of red and green channels; right, yellow (red + green) colocalization channel. Arrows denote yellow puncta that correspond to '1' and '2' in electron microscopy images below. Bottom left, electron microscopy of identical cell; middle and right, high magnification images of insets '1' and '2'. Scale bars, left, 10 μ m; middle and right, 200 nm. **b**, Ranked distribution of median Z-scores for each siRNA pool in primary colocalization (virophagy) screen. Red, decreased colocalization; green, increased colocalization; blue, insufficient numbers of green or red puncta per cell or total number of cells per well for analysis. **c**, Maps of protein interactions in enriched network modules (see Supplementary Fig. 3b). **d**, Association of siRNA hits with autophagy network.

forms of selective autophagy², including SIN capsid targeting to autophagosomes⁶.

Selective SIN autophagy (virophagy) promotes the survival of SIN-infected cells⁶. To determine if our identified candidate virophagy

genes have a similar function, we screened our confirmation siRNA library for genes that decreased cell survival after SIN infection. Two or more siRNAs targeting 98 of the genes decreased cell survival after SIN infection (Fig. 2, Supplementary Tables 3 and 6 and Supplementary Figs 1b and 2a); colocalization and cell survival effects of individual siRNAs were significantly correlated (Supplementary Fig. 2b) ($P = 3.8 \times 10^{-8}$, Spearman correlation). This is consistent with a pro-survival function of autophagic targeting of SIN capsid in virally infected cells⁶.

To investigate whether the identified candidate virophagy genes also function in other forms of selective autophagy, we performed a secondary screen for autophagy of damaged mitochondria (mitophagy). We used HeLa cells that express an mCherry fusion of Parkin, a cytosolic E3 ubiquitin ligase that translocates to depolarized mitochondria to induce mitophagy after treatment with uncoupling agents (such as CCCP, carbonyl cyanide *m*-chlorophenylhydrazone)¹³. Of the 141 confirmed colocalization hits, 2 or more siRNAs targeting 96 (68%) genes decreased mitophagy (Fig. 2, Supplementary Tables 3 and 7 and Supplementary Figs 1b and 2a, b). Host factors involved in viral autophagy and mitophagy overlapped significantly ($P = 0.019$, Spearman correlation). The minority of genes that only scored positive in either the virophagy confirmation or the mitophagy secondary screen may have a role in targeting some, but not other cargoes, for selective autophagy; however, the lack of overlap may also reflect different sensitivities of the two screens. Mitophagy hits consisted of several mitochondria-associated components¹⁴ (NME2, MDH1, NTHL1, PDK1, COX8A, MRPS2, MRPS10, NDUFB9 and BLOC1S) and interactors of mitochondria-associated components (Supplementary Fig. 4).

We focused further on one gene, *SMURF1* (SMAD specific E3 ubiquitin protein ligase 1), encoding a HECT-domain ubiquitin ligase that targets several cytoplasmic proteins for degradation¹⁵. *SMURF1* was a confirmed hit in all three confirmation or secondary screening assays (see Supplementary Fig. 5 for representative raw data from colocalization confirmation screen), is present in two of the enriched networks (mRNA processing and actin cytoskeleton) (Fig. 1c), and is a predicted interacting partner of the autophagy adaptor, p62 (refs 2, 4, Fig. 1d).

We confirmed that *SMURF1* is not required for general autophagy, but is a bona fide mediator of selective autophagy, including virophagy and mitophagy. siRNA knockdown of *SMURF1* in HeLa cells, unlike knockdown of the essential autophagy protein, ATG7, did not alter general starvation-induced autophagy (Supplementary Fig. 6a). Furthermore, *Smurf1*^{-/-} murine embryonic fibroblasts (MEFs) had normal levels of starvation-induced LC3-II (lipidated form of MAP1LC3) conversion, p62 degradation, and ultrastructural evidence of autophagosome and autolysosome accumulation (Fig. 3a, b).

However, a significant decrease in SIN/mCherry.capsid/GFP-LC3 colocalization was observed in SIN-infected *Smurf1*^{-/-} MEFs (Fig. 3c, d). Similar to p62 (ref. 6), *SMURF1* and SIN capsid protein co-immunoprecipitate in SIN-infected MEFs and HeLa cells (Supplementary Fig. 7a, b). *SMURF1* is not required for the interaction between p62 and SIN capsid (Supplementary Fig. 7c). The interaction between *SMURF1* and SIN capsid may be relevant for targeting SIN capsid for autophagosomal degradation, as levels of SIN capsid were increased in *Smurf1*^{-/-} MEFs and *SMURF1* siRNA-treated HeLa cells. (Supplementary Fig. 7d–f). Increased SIN capsid levels in *Smurf1*-deficient cells cannot be explained by increased capsid production because viral growth was similar in *Smurf1*^{-/-} and wild-type MEFs (Supplementary Fig. 7g, h), or by changes in proteasomal degradation because SIN capsid levels were not altered by treatment with the proteasome inhibitor MG132 (Supplementary Fig. 7d), and SIN capsid ubiquitination was not detected (data not shown). SIN-infected *Smurf1*^{-/-} MEFs had accelerated cell death (despite similar viral titres) as compared to wild-type controls (Supplementary Fig. 7i). Thus, *SMURF1* interacts with SIN capsid, *SMURF1* is required for

Gene symbol	C	S	M	Gene symbol	C	S	M	Gene symbol	C	S	M	Gene symbol	C	S	M	Gene symbol	C	S	M	Gene symbol	C	S	M	Gene symbol	C	S	M	Gene symbol	C	S	M
ACIN1	4	4	3	CALCB	0	1	3	EIF2AK1	2	0	1	HIST1H3H	3	2	0	MAP1A	3	2	4	NR3C1	0	0	1	RGS17	1	1	1	STK32A	4	0	2
ACTR1T	2	2	4	CAPS	2	0	2	EIF2S1	0	2	4	HPR	4	3	0	MAP2K1	3	0	4	NTLH1	2	3	4	RIMS3	2	0	4	STOM	2	4	2
ADAMTS1	3	1	3	CCDC36	4	3	1	ENCT	2	0	1	HSF2BP	2	3	0	MAP3K12	3	0	4	NTN4	2	1	1	SATB1	3	1	0	STX10	3	1	1
ADRB2	1	0	2	CD163L1	2	4	3	FABP1	0	4	2	IL13RA1	2	0	1	MBD5	4	4	2	NUP93	4	3	3	SCN1A	1	2	1	SUP3H	3	2	3
AKR1E2	3	1	3	CD93	2	3	3	FAM131B	4	2	3	IMP3	3	1	0	MDH1	4	3	3	OBSCN	3	2	2	SCRN1	1	0	1	TBC1D5	1	1	4
ALKBH5	4	4	1	CDK2AP1	1	1	1	FAM13B	2	2	4	IPPK	3	0	4	MEX3C	2	2	2	P2RX5	2	2	3	SERPINE1	3	4	2	TEAD4	1	3	1
ALPK1	3	3	3	CETN1	2	1	0	FAM176B	1	0	2	ITPKC	1	2	3	MLL3	3	1	1	P2RY4	1	0	0	SFRP4	4	0	3	TMEM203	2	4	2
ANXA5	3	3	2	CHAF1B	1	4	3	FANCC	3	0	3	KCNAB2	4	0	1	MRPS10	4	4	2	PDK1	1	1	4	SFRS4	2	0	1	TMEM39A	0	2	4
ASB2	3	4	4	CHCHD8	2	1	1	FANCF	4	1	2	KCNH3	0	0	1	MRPS2	3	2	3	PDK4	4	2	1	SLC1A3	2	4	3	TMEM39B	3	3	4
ATG13	2	1	2	CHST3	2	2	3	FANCL	2	4	1	KCNK3	2	0	4	MS4A4A	1	0	1	PEX13	2	3	3	SLC1A4	3	0	2	TNIK	2	2	0
ATP1B1	3	4	4	CLDN7	2	2	1	FCGR3B	1	2	4	KCNQ1	2	2	1	MSTN	3	3	3	PEX3	2	4	3	SLC22A3	1	0	3	TREM1	1	4	2
BET1	2	0	0	CLINT1	2	0	1	FGF14	3	4	0	KDR	3	1	0	MTSS1	1	2	1	PFKP	2	1	3	SLC25A19	3	3	0	TRPC5	3	0	0
BLOC1S1	0	0	2	CLVS1	1	1	2	FGF7	2	2	1	KIAA0174	1	3	2	MYH11	0	0	2	PGK2	2	1	4	SLC25A36	3	1	1	TXLNA	2	1	2
BMP2K1	4	2	2	CNOT7	1	0	1	FGFBP1	2	3	3	KIAA0232	0	0	0	MYLK	3	2	3	PHYHIP	3	3	3	SLC35B3	3	3	4	UBA52	2	2	1
BOC	3	1	2	COX6B1	1	1	1	FLJ25363	2	0	2	KRCC1	2	1	4	MYLK3	3	1	0	PI4KA	2	1	2	SLC35C1	3	4	4	YIPF1	2	3	3
C11orf41	3	3	3	COX8A	0	3	4	FXR2	2	2	1	KRT15	1	1	2	MYOM1	4	3	4	PIK3CA	1	0	4	SLC37A4	3	2	3	ZCCHC17	2	3	3
C14orf104	4	3	2	CPA3	4	2	1	GABRA5	3	3	2	KRT6A	3	3	4	NADSYN1	2	0	1	PLD2	1	1	3	SLC6A1	1	0	4	ZFYVE16	2	1	1
C1orf210	4	3	0	CRINK1	2	3	3	GAS2	2	0	0	KRT73	2	1	3	NAGA	0	0	1	PNPO	2	4	4	SLC10A2	1	1	2	ZNF189	1	3	2
C1orf223	2	4	2	CSPG5	4	4	2	GDF5	1	2	1	LARP1B	2	1	3	NDUFA4L2	1	1	0	PPY	3	4	2	SMURF1	2	2	2	ZNF593	2	1	4
C2orf12	1	1	2	CXCR7	0	4	1	GMIP	2	2	2	LENG9	3	3	3	NDUFB9	3	2	2	PRKD2	4	0	2	SNRNP	4	3	3	ZNF681	3	2	1
C3orf72	2	1	1	DAAM2	2	0	1	GNE	1	0	0	LMCD1	4	2	4	NEFM	3	0	2	PRKG1	2	0	2	SNRNP2	3	3	0				
C5	3	1	4	DHX38	1	1	0	GPC1	1	0	1	LOC387849	2	3	4	NF2	3	1	1	RBM18	1	3	0	SNRNP1	3	2	4				
C7orf68	2	2	1	DKKL1	4	3	2	GPR81	3	3	2	LOC402210	2	0	0	NLRP14	1	1	1	RCTP1	1	0	0	SNRPF	4	0	2				
C8orf59	3	2	2	DPF3	2	3	4	HAPLN1	1	1	2	LRPAP1	1	0	0	NME2	1	1	4	REEP2	4	2	1	SNTG1	0	2	2				
CA7	2	2	3	DUSP22	0	1	0	HEATR6	3	1	1	LSM4	4	4	3	NR2C2	2	1	4	RFWD3	4	3	2	STAT2	3	4	2				

Figure 2 | Gene list for viral capsid/autophagosome colocalization (C) confirmation screen and secondary screens for survival of virus-infected cells (S) and Parkin-mediated mitophagy (M). Shown are the numbers of individual siRNAs from a pool of four targeting each gene that scored positive

in each screen. Red, genes with 2 or more positive siRNAs (confirmed hits); Green, genes with <2 positive siRNAs. See Supplementary Tables 2–8 for further details.

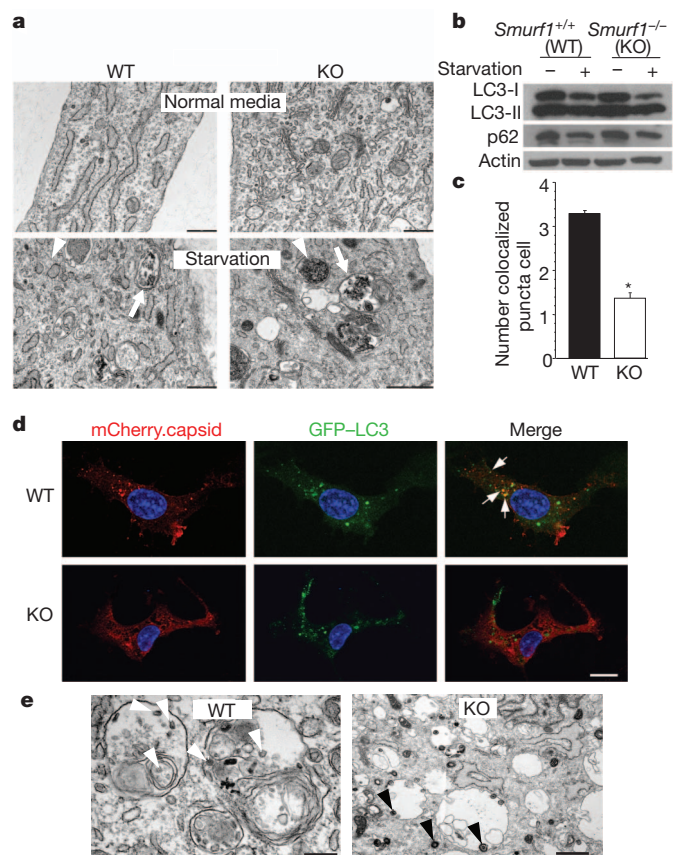


Figure 3 | SMURF1 functions in virophagy but not in starvation-induced autophagy. **a**, Electron microscopy analysis of wild-type (WT) and *Smurf1*^{-/-} (KO) MEFs grown in normal media or EBSS (starvation) for 4 h. Arrowheads, representative autophagosomes; arrows, representative autolysosomes. Scale bars, 500 nm. **b**, Western blot analyses of LC3-I/II (non-lipidated and lipidated forms of MAP1LC3, respectively) and p62 levels in MEFs of indicated genotype. **c**, Quantification of colocalization of SIN/mCherry.capsid and GFP-LC3 in indicated MEFs 16 h after SIN/mCherry.capsid/GFP-LC3 infection. Data shown represent mean \pm s.e.m. number of colocalized red and green puncta per cell for 50 cells per well in triplicate samples. * P < 0.001 against wild-type, Student's t -test. **d**, Representative confocal microscopy images of images used for quantification in **c**. Arrows, colocalized red and green puncta. Scale bar, 15 μ m. **e**, Representative electron microscopy images of indicated MEFs infected with HSV-1 (strain 17termA). White arrowheads, partially degraded viral nucleocapsids inside autolysosomes; black arrowheads, intact viral nucleocapsids inside viral vesicles. Scale bar, 0.5 μ m. For **a–e**, similar results were obtained in 3–5 independent experiments.

SIN capsid targeting to autophagosomes and degradation through a proteasome-independent pathway, and SMURF1-dependent degradation of SIN capsid promotes cell survival.

To determine whether SMURF1 is required for the autophagic targeting of other viruses, we performed electron microscopy of wild-type and *Smurf1*^{-/-} MEFs infected with a mutant strain of herpes simplex virus type 1 harbouring a deletion of ICP34.5, a potent inhibitor of viral autophagy^{16–18} (Fig. 3e). As reported¹⁷, the majority of cytoplasmic HSV-1 virions in wild-type MEFs were inside autolysosomal structures and appeared partially degraded. In contrast, in *Smurf1*^{-/-} MEFs, the majority of cytoplasmic HSV-1 virions were inside single-membraned vesicles involved in HSV-1 cytoplasmic egress and had an intact structure. This lack of autophagic targeting of HSV-1 in *Smurf1*^{-/-} MEFs was not due to a general defect in autophagy, because HSV-1 infection induced autophagy similarly in *Smurf1*^{-/-} and wild-type MEFs (Supplementary Fig. 6b). Thus, SMURF1 is required for the autophagic targeting of both a positive-strand RNA (Sindbis) and a double-stranded DNA (herpes simplex) virus.

Next, we examined the role of SMURF1 in mitophagy. In HeLa cells, all four *SMURF1* siRNAs decreased SMURF1 protein expression (Supplementary Fig. 8a) and inhibited Parkin-mediated CCCP-induced mitophagy as effectively as an siRNA targeting *p62*, a mediator of mitophagy in some previous reports^{19,20}, and siRNA targeted against the essential autophagy gene, *ATG7* (Supplementary Fig. 8b, c). The magnitude of each individual siRNA's effect on SMURF1 protein expression knockdown correlated with the magnitude of inhibition of Parkin-mediated autophagy. Therefore, in the mitophagy confirmation screen, two of the *SMURF1* siRNAs were probably false negatives; indeed, the number of Parkin-expressing cells in wells treated with these siRNAs was low (data not shown), precluding meaningful statistical analyses. A similar finding was true in the viral colocalization screen.

We further examined the role of SMURF1 in mitophagy by assessing mitochondrial clearance in CCCP-treated *Smurf1*^{-/-} MEFs. Unlike in HeLa cells, Parkin overexpression did not promote mitophagy in MEFs of either genotype (data not shown). However, 25–30% of MEFs treated with 10 μ M CCCP showed changes in mitochondrial morphology. In wild-type MEFs with damaged mitochondria (swollen or fragmented appearance), partial mitochondrial clearance occurred with compaction of the remaining mitochondria around the nucleus (Fig. 4a). In contrast, in *Smurf1*^{-/-} cells with damaged mitochondria, virtually no mitochondrial clearance occurred and there was diffuse accumulation of fragmented mitochondria throughout the cytoplasm (Fig. 4a, arrows). This phenotypic difference was confirmed using two independent methods of quantification, including assessment of the total percentage of CCCP-treated cells that displayed diffuse

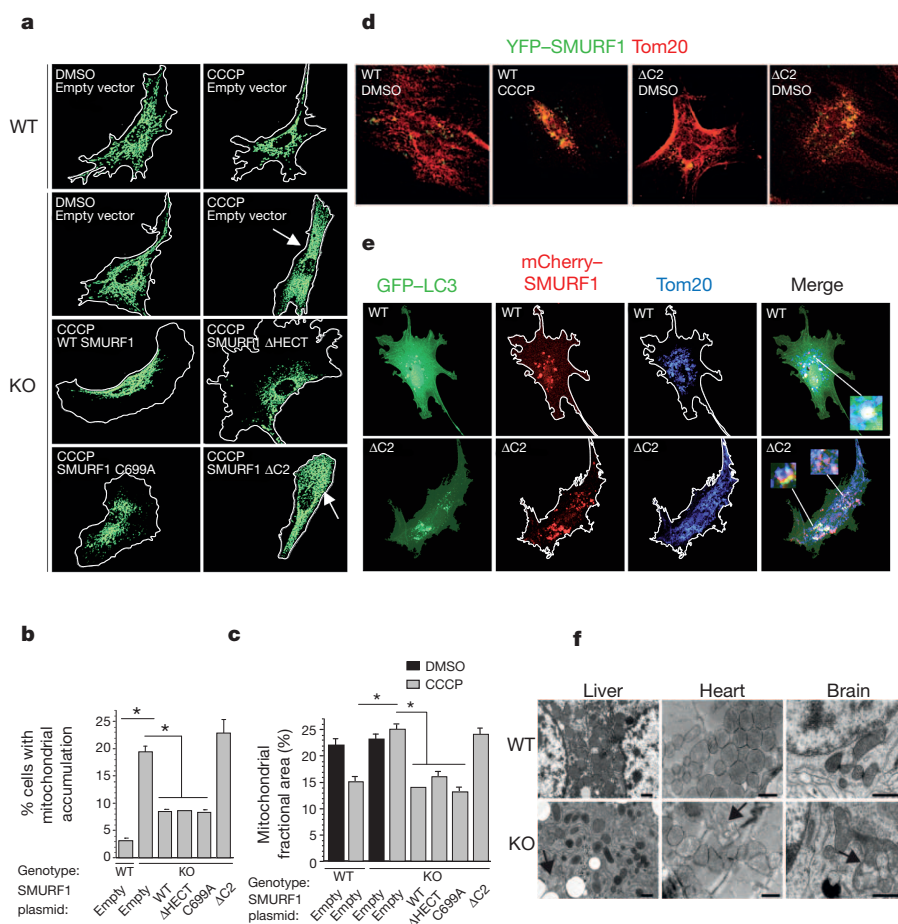


Figure 4 | SMURF1 functions in mitophagy.

a, Representative mitochondrial morphology in *Smurf1*^{+/+} (wild-type) and *Smurf1*^{-/-} (KO) MEFs transfected with indicated construct and treated with DMSO or 10 μM CCCP for 24 h.

b, Quantification of percentage of total cells with a diffuse accumulation of abnormal fragmented mitochondria and lack of mitochondrial clearance. Results shown represent combined data from 3–5 experiments per condition with triplicate wells (of at least 100 cells per well) analysed for each condition per experiment. Shown are mean ± s.e.m. for average values from each experiment. Similar results were observed in each independent experiment. **P* < 0.001, Student's *t*-test. **c**, Measurement of mitochondrial fractional area (percentage of total cellular area) in MEFs treated as in **a**. Results shown represent mean ± s.e.m. for 50 cells per condition.

d, Representative confocal micrographs of KO MEFs transfected with YFP-SMURF1 wild-type or YFP-SMURF1ΔC2 (ΔC2) and treated for 4 h with DMSO or CCCP. **e**, Representative confocal micrographs of KO MEFs transfected with GFP-LC3 and wild-type mCherry-SMURF1 (WT) or mCherry-SMURF1ΔC2 (ΔC2) and treated for 4 h with CCCP. Inset, upper right, formation of completed autophagosome around a damaged mitochondrion associated with wild-type SMURF1; insets, lower right, incomplete autophagosomes or absence of LC3 signal around mitochondria associated with SMURF1ΔC2. See also Supplementary Figs 10 and 11 for enlarged images. **f**, Representative electron microscopy images of indicated tissues from 10-month-old *Smurf1*^{+/+} (WT) or *Smurf1*^{-/-} (KO) mice.

Arrows, representative abnormal mitochondria. Scale bars, 1 μm. Similar abnormalities were observed throughout entire electron microscopy tissue section for each mouse and in tissue samples from three different mice for each genotype.

accumulation of abnormal mitochondria (Fig. 4b) and the measurement of fractional mitochondrial surface area per cell (Fig. 4c).

To evaluate the mechanism of action of SMURF1, we compared the effects of wild-type and mutant SMURF1 expression plasmids on rescue of selective autophagy in *Smurf1*^{-/-} MEFs (Fig. 4a–c). We focused on mitophagy rather than SIN capsid viroplasm because of the resistance of MEFs to SIN infection after plasmid transfection. The defect in mitophagy in *Smurf1*^{-/-} MEFs was partially rescued by wild-type SMURF1 transfection. SMURF1ΔHECT²¹, lacking the HECT domain that catalyses ubiquitin ligation onto target proteins, or SMURF1(C699A)²¹, a catalytically inactive point mutant, rescued the mitochondrial clearance defect as efficiently as wild-type SMURF1. Thus, in addition to its known role in targeting proteins for proteasomal degradation via ubiquitination, SMURF1 has a ubiquitin ligase activity-independent function in mediating the selective degradation of damaged mitochondria.

In contrast, a SMURF1 mutant lacking the C2 domain, SMURF1ΔC2, was completely defective in mitophagy rescue in CCCP-treated *Smurf1*^{-/-} MEFs (Fig. 4a–c), despite similar levels of expression as transfected wild-type SMURF1 (Supplementary Fig. 9a). The C2 domain of SMURF1 was not required for SMURF1 co-immunoprecipitation with p62 (Supplementary Fig. 9b), indicating that SMURF1 does not function in selective autophagy by recruiting p62. C2 domains (including those of protein kinase C and SMURF1) bind membrane phospholipids and function in protein targeting to the plasma membrane and/or membrane subcellular compartments^{22,23}. This raised the possibility that SMURF1 might function in the targeting of selective autophagy cargo through interaction with the nascent autophagosome membrane.

To investigate this possibility, we examined the subcellular localization of wild-type SMURF1 and SMURF1ΔC2 with damaged mitochondria and autophagosomes (Fig. 4d). In *Smurf1*^{-/-} MEFs transfected with wild-type, yellow fluorescent protein-conjugated YFP-SMURF1, CCCP treatment induced the colocalization of YFP-SMURF1 with damaged mitochondria. In *Smurf1*^{-/-} MEFs transfected with YFP-SMURF1ΔC2, increased numbers of fragmented and swollen mitochondria were observed in basal conditions and these increased further upon CCCP treatment. These abnormal mitochondria colocalized with YFP-SMURF1ΔC2, whereas normal reticular-appearing mitochondria rarely colocalized with YFP-SMURF1ΔC2. YFP-SMURF1(C699A) displayed the same subcellular staining pattern as wild-type YFP-SMURF1 (data not shown). Thus, SMURF1 colocalizes with damaged mitochondria in a C2 domain-independent manner.

We next determined whether the C2 domain of SMURF1 was required for the colocalization of damaged mitochondria with autophagosomes (Fig. 4e and Supplementary Figs 10 and 11). In cells expressing wild-type mCherry-SMURF1, mitochondria were mostly compacted around the nucleus, and numerous autophagosomes were observed surrounding structures that labelled positive for both mCherry-SMURF1 and the mitochondrial marker, Tom20 (also known as TOMM20). In contrast, in cells expressing mCherry-SMURF1ΔC2, mCherry-SMURF1ΔC2- and Tom20-positive mitochondria were rarely found inside autophagosomes. In many regions, GFP-LC3-positive linear or cup-shaped structures were observed near mCherry-SMURF1ΔC2-positive mitochondria, but complete autophagosomes surrounding these mitochondria could not be detected. Thus, the C2 domain of SMURF1 is

not required for its targeting to damaged mitochondria, but is required for damaged mitochondria to be normally engulfed by autophagosomes. It is not yet known whether this requirement reflects a direct role for the C2 domain in binding to autophagosomal membrane phospholipids or is a more indirect consequence of other, as-yet-undescribed, effects of the C2 domain in mitophagy.

To investigate whether SMURF1 may function in selective autophagy *in vivo*, we performed electron microscopy analyses of cerebellum, liver and hearts of 10-month-old wild-type and *Smurf1*^{-/-} mice²¹. In all three organs, *Smurf1*^{-/-} mice showed an accumulation of abnormal mitochondria that were swollen, fragmented, and/or contained abnormal cristae (Fig. 4f). This phenotype is consistent with a defect in mitophagy and mitochondrial quality control; however, we cannot rule out unknown triggers of mitochondrial damage in these animals. In the livers of *Smurf1*^{-/-} mice, mitochondria were spatially disorganized and surrounded by networks of dilated endoplasmic reticulum, perhaps reflecting a defect in mitochondrial targeting by isolation membranes (which are believed to originate from the endoplasmic reticulum¹) and/or a defect in selective autophagy of the endoplasmic reticulum. There was a marked accumulation of lipid droplets in the livers of *Smurf1*^{-/-} mice (Supplementary Fig. 12a), which may be consistent with selective degradation of lipid droplets by autophagy (lipophagy) in hepatocytes²⁴. Furthermore, the granule cell layer of the cerebellum and cardiomyocytes of *Smurf1*^{-/-} mice had increased numbers of p62 aggregates (Supplementary Fig. 12b). Unlike findings in brains and hearts of mice lacking core autophagy genes²⁵, p62 aggregate accumulation in these tissues was not associated with ubiquitin accumulation. This is consistent with a role for SMURF1 in selective autophagy but not in the form of basal autophagy that is involved in protein quality control²⁵.

Together, our data in *Smurf1*^{-/-} MEFs and in *Smurf1*^{-/-} mice suggest a crucial function for SMURF1 in selective autophagy, including in the autophagic targeting of genetically distinct viruses, in the autophagic targeting of mitochondria and, more speculatively, in the potential autophagic targeting of other cellular targets such as hepatic lipid droplets and endoplasmic reticulum. The mechanism by which SMURF1 functions in selective autophagy is independent of its E3 ubiquitin ligase activity, but rather involves its C2 membrane-targeting domain. We propose that the C2 domain of SMURF1 may participate in the delivery of selective autophagic substrates to the nascent autophagosome. Thus, SMURF1 has parallel functions in two distinct cellular degradation pathways, targeting specific proteins for degradation by the ubiquitin-proteasomal pathway¹⁵ (via its E3 ubiquitin ligase activity) and targeting selective cargo for degradation by the autophagy pathway (via its C2 domain).

Our findings in *Smurf1*^{-/-} MEFs and mice illustrate that our high-content image-based genome-wide screen successfully reveals novel candidate determinants of selective autophagy. More broadly, the identification of a set of 96 genes that may dually function in viral autophagy and mitophagy (but not in basal autophagy) suggests the existence of a common molecular network for targeting diverse unwanted cytoplasmic cargo to the lysosome. This network identification provides a basis for a more global understanding of the mechanisms involved in selective autophagy.

METHODS SUMMARY

High-content image-based genome-wide siRNA screen. A genome-wide siRNA library (Dharmacon) containing 21,125 SMART pools was used for reverse transfection of HeLa/GFP-LC3 cells, followed by infection with SIN/mCherry.capsid virus, high content imaging using a Pathway855 automated microscope (BD Biosciences), quantitative image analysis, statistical analysis, and bioinformatic analysis as described in Supplementary Information. Primary hits were evaluated in three confirmation/secondary screens using the four individual siRNAs from each pool, including a screen for viral capsid/autophagosome colocalization, cell survival during SIN infection, and Parkin-induced mitophagy.

Functional analyses of SMURF1. See Supplementary Information.

Received 6 September 2010; accepted 7 September 2011.

Published online 23 October; corrected 1 December 2011 (see full-text HTML version for details).

- Levine, B., Mizushima, N. & Virgin, H. W. Autophagy in immunity and inflammation. *Nature* **469**, 323–335 (2011).
- Noda, N. N., Ohsumi, Y. & Inagaki, F. Atg8-family interacting motif crucial for selective autophagy. *FEBS Lett.* **584**, 1379–1385 (2010).
- Wild, P. *et al.* Phosphorylation of the autophagy receptor optineurin restricts *Salmonella* growth. *Science* **333**, 228–233 (2011).
- Komatsu, M. & Ichimura, Y. Selective autophagy regulates various cellular functions. *Genes Cells* **15**, 923–933 (2010).
- Lipinski, M. M. *et al.* A genome-wide siRNA screen reveals multiple mTORC1 independent signaling pathways regulating autophagy under normal nutritional conditions. *Dev. Cell* **18**, 1041–1052 (2010).
- Orvedahl, A. O. *et al.* Autophagy protects against Sindbis virus infection of the central nervous system. *Cell Host Microbe* **7**, 115–127 (2010).
- Monastyrska, I., Rieter, E., Klionsky, D. J. & Reggiori, F. Multiple roles of the cytoskeleton in autophagy. *Biol. Rev. Camb. Philos. Soc.* **84**, 431–448 (2009).
- Lee, J. Y. *et al.* HDAC6 controls autophagosome maturation essential for ubiquitin-selective quality-control autophagy. *EMBO J.* **29**, 969–980 (2010).
- Longatti, A. & Tooze, S. A. Vesicular trafficking and autophagosome formation. *Cell Death Differ.* **16**, 956–965 (2009).
- Nair, U. *et al.* SNARE proteins are required for macroautophagy. *Cell* **146**, 290–302 (2011).
- Behrends, C., Sowa, M. E., Gygi, S. P. & Harper, J. W. Network organization of the human autophagy system. *Nature* **466**, 68–76 (2010).
- Mizushima, N. The role of the Atg1/ULK1 complex in autophagy regulation. *Curr. Opin. Cell Biol.* **22**, 132–139 (2010).
- Narendra, D., Tanaka, A., Suen, D. F. & Youle, R. J. Parkin is recruited selectively to impaired mitochondria and promotes their autophagy. *J. Cell Biol.* **183**, 795–803 (2008).
- Pagliarini, D. J. *et al.* A mitochondrial protein compendium elucidates complex I disease biology. *Cell* **134**, 112–123 (2008).
- Xing, L., Zhang, M. & Chen, D. Smurf control in bone cells. *J. Cell. Biochem.* **110**, 554–563 (2010).
- Talloczy, Z. *et al.* Regulation of starvation- and virus-induced autophagy by the eIF2 α kinase signaling pathway. *Proc. Natl Acad. Sci. USA* **99**, 190–195 (2002).
- Talloczy, Z., Virgin, H. W. I. V. & Levine, B. PKR-dependent xenophagic degradation of herpes simplex virus type 1. *Autophagy* **2**, 24–29 (2006).
- Orvedahl, A. *et al.* HSV-1 ICP34.5 confers neurovirulence by targeting the Beclin 1 autophagy protein. *Cell Host Microbe* **1**, 23–35 (2007).
- Geisler, S. *et al.* PINK1/Parkin-mediated mitophagy is dependent on VDAC1 and p62/SQSTM1. *Nature Cell Biol.* **12**, 119–131 (2010).
- Narendra, D., Kane, L. A., Hauser, D. N., Fearnley, I. M. & Youle, R. J. p62/SQSTM1 is required for Parkin-induced mitochondrial clustering but not mitophagy; VDAC1 is dispensable for both. *Autophagy* **6**, 1090–1106 (2010).
- Yamashita, M. *et al.* Ubiquitin ligase Smurf1 controls osteoblast activity and bone homeostasis by targeting MEKK2 for degradation. *Cell* **121**, 101–113 (2005).
- Cho, W. & Stahelin, R. V. Membrane binding and subcellular targeting of C2 domains. *Biochim. Biophys. Acta* **1761**, 838–849 (2006).
- Lu, K. *et al.* Pivotal role of the C2 domain of the Smurf1 ubiquitin ligase in substrate selection. *J. Biol. Chem.* **286**, 16861–16870 (2011).
- Singh, R. *et al.* Autophagy regulates lipid metabolism. *Nature* **458**, 1131–1135 (2009).
- Mizushima, N. & Levine, B. Autophagy in mammalian development and differentiation. *Nature Cell Biol.* **12**, 823–830 (2010).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank M. Vishwanath, S. Wei and B. Posner for assistance with high-throughput siRNA screening; W. Sun for information technology support; K. Scudder for assistance with image analysis algorithms; A. Diehl for expert medical illustration; V. Stollar, M. McDonald, R. Kuhn and R. Youle for helpful discussions and providing reagents; A. Bugde for assistance in the UTSW Live Cell Imaging Facility; and L. Mueller and T. Januszewski for assistance with electron microscopy. This work was supported by NIH grants AI109617 (B.L.), CA84254 (B.L.), UL1 RR024982 (G.X., Y.X.), AI062773 (R.J.X.), DK83756 (R.J.X.), DK086502 (R.J.X.) and DK043351 (R.J.X. and A.N.); NSF grant DMS-0907562 (G.X.); and the Center for Cancer Research, National Cancer Institute Intramural Research Program (Y.E.Z.).

Author Contributions A.O., R.S., M.N., M.R., J.L.W., Y.E.Z., K.L.-P., C.G. and B.L. designed the experiments. A.O., R.S., Z.Z. Q.S. and Y.T. performed the experiments. G.X., A.N., C.V.F., R.J.X. and Y.X. performed statistical and bioinformatic analyses. A.O., R.S. and B.L. wrote the manuscript. G.X. and A.N. contributed equally to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.L. (beth.levine@utsouthwestern.edu).

Nuclear PKM2 regulates β -catenin transactivation upon EGFR activation

Weiwei Yang¹, Yan Xia¹, Haitao Ji¹, Yanhua Zheng¹, Ji Liang¹, Wenhua Huang², Xiang Gao³, Kenneth Aldape⁴ & Zhimin Lu^{1,5,6}

The embryonic pyruvate kinase M2 (PKM2) isoform is highly expressed in human cancer. In contrast to the established role of PKM2 in aerobic glycolysis or the Warburg effect^{1–3}, its non-metabolic functions remain elusive. Here we demonstrate, in human cancer cells, that epidermal growth factor receptor (EGFR) activation induces translocation of PKM2, but not PKM1, into the nucleus, where K433 of PKM2 binds to c-Src-phosphorylated Y333 of β -catenin. This interaction is required for both proteins to be recruited to the *CCND1* promoter, leading to HDAC3 removal from the promoter, histone H3 acetylation and cyclin D1 expression. PKM2-dependent β -catenin transactivation is instrumental in EGFR-promoted tumour cell proliferation and brain tumour development. In addition, positive correlations have been identified between c-Src activity, β -catenin Y333 phosphorylation and PKM2 nuclear accumulation in human glioblastoma specimens. Furthermore, levels of β -catenin phosphorylation and nuclear PKM2 have been correlated with grades of glioma malignancy and prognosis. These findings reveal that EGF induces β -catenin transactivation via a mechanism distinct from that induced by Wnt/Wingless⁴ and highlight the essential non-metabolic functions of PKM2 in EGFR-promoted β -catenin transactivation, cell proliferation and tumorigenesis.

Because both EGFR activation and PKM2 expression are instrumental in tumorigenesis^{5–7}, we examined whether EGFR activation regulates PKM2 functions in a subcellular-compartment-dependent manner. Immunofluorescence analysis showed that EGF treatment resulted in the nuclear accumulation of PKM2 in U87/EGFR human glioblastoma (GBM) cells (Fig. 1a). In addition, expression of the constitutively active EGFRvIII mutant in U87 cells resulted in a higher amount of nuclear PKM2 than was observed in EGF-untreated U87/EGFR cells (Supplementary Fig. 2a). The finding that EGF induces nuclear translocation of PKM2 was further supported by cell fractionation analysis of DU145 prostate cancer cells, MDA-MB-231 breast cancer cells and U87/EGFR cells (Supplementary Fig. 2b). In addition, PKM1 failed to translocate into the nucleus upon EGF stimulation (Supplementary Fig. 2c), indicating that EGF specifically regulates the subcellular distribution of PKM2 in multiple types of cancer cells.

To examine whether PKM2 directly regulates gene transcription and cell proliferation, we expressed PKM2 short hairpin RNA (shRNA) in U87/EGFR cells (Supplementary Fig. 3a). PKM2 depletion largely reduced both basal and EGF-induced tumour cell proliferation (Fig. 1b) and blocked EGF-enhanced expression of cyclin D1 and c-Myc (Fig. 1c), which are known to be important regulators of cell proliferation and downstream genes of β -catenin transactivation⁸. To examine whether these PKM2-dependent effects were mediated by β -catenin, we performed TCF/LEF-1 luciferase reporter analyses, showing that PKM2 depletion significantly inhibited EGF-induced β -catenin transactivation (Fig. 1d). In addition, chromatin immunoprecipitation (ChIP) analyses showed that EGFR activation resulted in increased binding of β -catenin to the promoter region of *CCND1* (coding for cyclin

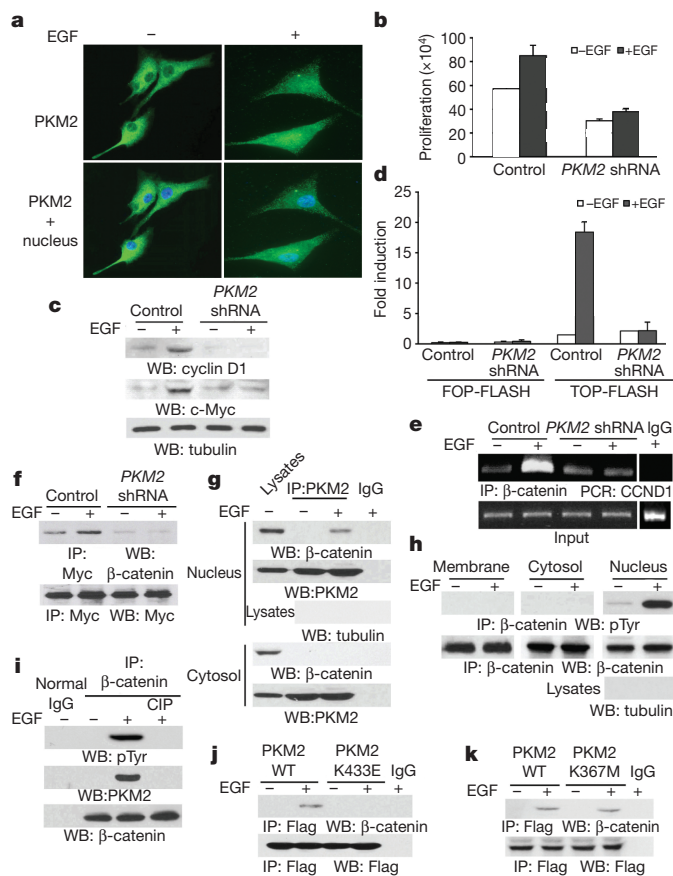


Figure 1 | EGF induces the PKM2- β -catenin interaction in the nucleus.

a, U87/EGFR cells were treated with or without EGF for 10 h. **b**, U87/EGFR cells with or without PKM2 depletion were plated and counted 7 days after seeding. Data represent the means \pm s.d. of three independent experiments. **c**, **e**, U87/EGFR cells with or without PKM2 depletion were treated with or without EGF for 24 h (**c**) or 10 h (**e**). IP, immunoprecipitation; WB, western blot. **d**, U87/EGFR cells with or without PKM2 depletion were transfected with TOP-FLASH or FOP-FLASH, which was followed by EGF treatment for 10 h. Data represent the means \pm s.d. of three independent experiments. **f**, Myc-TCF4 was immunoprecipitated from PKM2-depleted or PKM2-undepleted U87/EGFR cells treated with or without EGF for 10 h. **g**, **h**, PKM2 (**g**) or β -catenin (**h**) was immunoprecipitated from the indicated cell fractions of U87/EGFR cells treated with or without EGF for 6 h. **i**, β -Catenin immunoprecipitated from U87/EGFR cells with or without EGF treatment for 6 h was incubated with or without CIP (10 units) for 30 min at 37 °C followed by washing with PBS for three times. **j**, **k**, U87/EGFR cells stably expressing Flag-tagged wild-type PKM2, PKM2 K433E (**j**), or PKM2 K367M (**k**) were treated with or without EGF for 6 h.

¹Brain Tumor Center and Department of Neuro-Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ²Jiaxing Xinda Biotechnology Company, 1369 Cheng Nan Road, Science and Technology Building, STE 112, Jiaxing, Zhejiang Province 314000, China. ³Model Animal Research Center, State Key Laboratory of Pharmaceutical Biotechnology, Nanjing University, Nanjing 210061, China. ⁴Department of Pathology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ⁵Department of Molecular and Cellular Oncology, The University of Texas MD Anderson Cancer Center, Houston, Texas 77030, USA. ⁶The University of Texas Graduate School of Biomedical Sciences at Houston, Houston, Texas 77030, USA.

D1) (Fig. 1e) and *c-Myc* (data not shown), which was blocked by PKM2 depletion. In addition, co-immunoprecipitation analyses showed that PKM2 depletion inhibited the EGF-induced interaction between β -catenin and Myc-tagged TCF4 (Fig. 1f). However, PKM2 depletion failed to inhibit WNT3A- or WNT1- (data not shown) induced β -catenin transactivation (Supplementary Fig. 3b) and cyclin D1 expression (Supplementary Fig. 3c). In addition, WNT3A did not induce PKM2 nuclear translocation (Supplementary Fig. 3d). These results indicate that EGF induces β -catenin transactivation via a mechanism distinct from that induced by Wnt/Wingless⁴ and that PKM2 expression has a pivotal role in EGF-, but not Wnt-induced β -catenin transactivation.

To examine the mechanism underlying PKM2-regulated β -catenin transactivation, we performed co-immunoprecipitation analyses, showing that EGF stimulation resulted in an interaction between endogenous PKM2 and β -catenin in the nuclear, but not cytosolic, fraction of U87/EGFR cells (Fig. 1g). However, an *in vitro* glutathione S-transferase (GST) pull-down assay showed that purified GST- β -catenin failed to bind to purified His-PKM2 (Supplementary Fig. 4). These results suggest that the interaction of these two proteins might require post-translational modifications of the proteins.

PKM2 binds to tyrosine-phosphorylated peptides, and expression of the phosphotyrosine-binding form is required for cancer cell growth⁹. To examine whether β -catenin is tyrosine phosphorylated, we performed immunoblotting analyses with a phospho-Tyr antibody, showing that EGF stimulation induced Tyr phosphorylation of immunoprecipitated β -catenin in the nucleus, but not in the cytosol or at the plasma membrane (Fig. 1h). Treatment of the immunoprecipitated β -catenin with calf intestinal alkaline phosphatase (CIP) resulted in β -catenin dephosphorylation and abrogation of the PKM2- β -catenin interaction (Fig. 1i). Thus, EGF-induced Tyr phosphorylation of β -catenin is required for the PKM2- β -catenin interaction.

The PKM2 K433E mutant, which fails to bind to tyrosine-phosphorylated peptides⁹, had similar glycolytic enzyme activity to its wild-type counterpart (Supplementary Fig. 5)⁹. Co-immunoprecipitation analyses showed that EGF treatment induced the binding of β -catenin to Flag-tagged wild-type PKM2, but not to the PKM2 K433E mutant (Fig. 1j). In contrast, a kinase-dead Flag-PKM2 K367M mutant^{10,11}, acting like its wild-type counterpart, binds to β -catenin (Fig. 1k). These results indicate that the K433 binding residue of PKM2, but not its catalytic activity, is critical for the PKM2- β -catenin interaction.

Abl and c-Src have been reported to phosphorylate β -catenin^{12,13}. Pre-treatment with SU6656 (an Src inhibitor) or an Abl inhibitor completely abrogated EGF-induced activation of c-Src or Abl, as shown by the reduced levels of c-Src (Y418) or Abl (Y412) phosphorylation (Supplementary Fig. 6a). However, inhibition of c-Src, but not Abl, blocked EGF-induced Tyr phosphorylation of β -catenin (Fig. 2a). In addition, deficiency of c-Src (Fig. 2b), but not of Abl (Supplementary Fig. 6b), abrogated EGF-induced β -catenin Tyr phosphorylation and the PKM2- β -catenin interaction. These results indicate that c-Src, a downstream effector of EGFR, phosphorylates β -catenin, which is required for the PKM2- β -catenin interaction.

To examine the subcellular compartment in which c-Src phosphorylates β -catenin, we conducted fractionation analyses. We found that EGF stimulation resulted in the nuclear translocation of c-Src (Supplementary Fig. 6c). In addition, co-immunoprecipitation analyses showed that EGF treatment induced an enhanced interaction between β -catenin and c-Src in nuclear fractions (Fig. 2c). These results, combined with the evidence that β -catenin is phosphorylated in the nucleus (Fig. 1h), strongly suggest that c-Src translocates into the nucleus and subsequently interacts with and phosphorylates β -catenin.

The Y86 residue of β -catenin has been shown to be phosphorylated by c-Src as well as BCR-ABL^{12,13}, and analysis of the amino acid sequence identified an additional potential Src phosphorylation site at Y333. Immunoblotting analysis showed that EGF stimulation resulted in tyrosine phosphorylation of Flag-tagged wild-type β -catenin and β -catenin Y86F, but not β -catenin Y333F, which was further validated by immunoblotting with a phospho- β -catenin Y333 antibody (Fig. 2d). In addition, cell fractionation analysis demonstrated that EGF induced β -catenin Y333 phosphorylation primarily in the nucleus (Supplementary Fig. 6d). Furthermore, an *in vitro* protein kinase assay showed that active c-Src was able to phosphorylate wild-type β -catenin (Fig. 2e), but the β -catenin Y333F mutation largely reduced total Tyr-phosphorylation levels and completely abrogated the Y333 phosphorylation. These results reveal that c-Src binds and phosphorylates β -catenin at Y333 *in vitro* and *in vivo*.

To examine whether phosphorylation of β -catenin Y333 regulates its binding to PKM2, we performed a GST pull-down assay by mixing purified GST- β -catenin and His-PKM2 with or without purified active c-Src. Figure 2f shows that wild-type PKM2 did not bind to unphosphorylated wild-type β -catenin. However, the presence of c-Src enabled the binding of wild-type PKM2 to β -catenin, which was abrogated by mutation of β -catenin at Y333 or PKM2 at K433. These *in vitro* results were further validated by co-immunoprecipitation analyses, showing that in contrast to Flag-tagged wild-type β -catenin or β -catenin Y86F, Flag- β -catenin Y333F failed to bind to endogenous PKM2 (Fig. 2g). These results indicate that c-Src-mediated β -catenin Y333 phosphorylation is required for the PKM2- β -catenin interaction.

We next examined the importance of the PKM2- β -catenin interaction in β -catenin transactivation. Figure 3a shows that Flag- β -catenin Y333F failed to interact with Myc-tagged TCF4 upon EGF stimulation, in contrast to its wild-type counterpart. ChIP analyses demonstrated that Flag-tagged wild-type β -catenin bound to the *CCND1* promoter region upon EGFR activation, which was abrogated by the Y333F mutation (Fig. 3b). In addition, reconstituted expression of the RNA interference (RNAi)-resistant (r) β -catenin Y333F mutant, but not of wild-type r β -catenin, in endogenous β -catenin-depleted U87/EGFRvIII cells failed to induce cyclin D1 and c-Myc expression (Fig. 3c). Furthermore, Flag-PKM2 K433E and the inactive Flag-PKM2 K367M mutant translocated into the nucleus upon EGF stimulation (Supplementary Fig. 7), but reconstituted expression of these mutants (Supplementary Fig. 8) failed to induce cyclin D1 expression as did wild-type rPKM2 expression (Fig. 3d). Thus, both

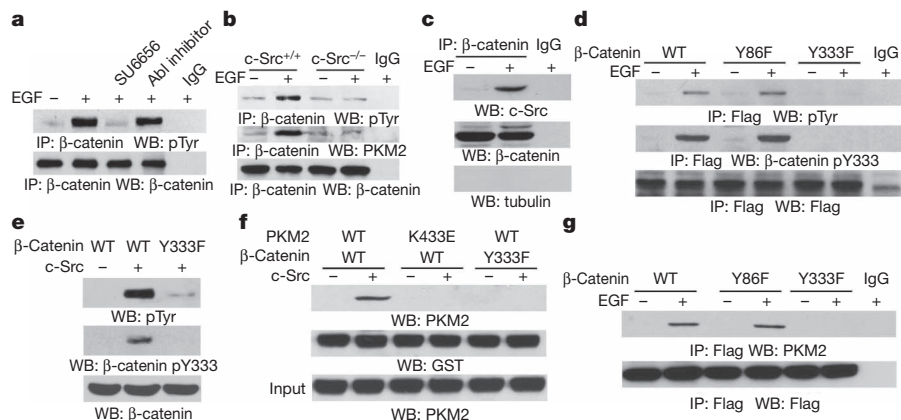


Figure 2 | c-Src phosphorylates β -catenin at Y333 upon EGFR activation. **a**, U87/EGFR cells were treated with SU6656 (4 μ M) or an Abl inhibitor (0.2 μ M) for 30 min before EGF treatment for 6 h. **b**, The indicated cells were treated with or without EGF for 6 h. **c**, β -Catenin was immunoprecipitated from the nuclear fractions of U87/EGFR cells treated with or without EGF for 6 h. **d**, **e**, U87/EGFR cells transiently expressing the indicated Flag-tagged β -catenin proteins were treated with or without EGF for 6 h. **e**, *In vitro* kinase assays were performed with purified active c-Src and purified β -catenin proteins. WT, wild type. **f**, Immobilized GST- β -catenin proteins were mixed with purified His-PKM2 proteins in the presence or absence of active c-Src.

PKM2 catalytic activity and the PKM2- β -catenin interaction are required for cyclin D1 expression upon EGFR activation.

To compare downstream targets of EGF and Wnt signalling, we examined the expression of other Wnt/ β -catenin downstream genes: *AXIN2*, *DKK1* and *β TrCP* (also known as *BTRCP*)⁸. Quantitative reverse transcription-polymerase chain reaction (RT-PCR) analysis showed that EGF treatment increased mRNA levels of *DKK1*, but not of *AXIN2* or *β TrCP*, which was blocked by PKM2 depletion (Supplementary Fig. 9). These results indicate that EGF-induced and PKM2-dependent β -catenin transactivation induced transcription of a set of genes, which do not completely overlap with those induced by Wnt signalling.

The finding that PKM2 was not required for WNT3A-induced β -catenin transactivation (Supplementary Fig. 3b) suggested that c-Src-dependent β -catenin Y333 phosphorylation is not involved in Wnt-induced signalling or cell adhesion. This assumption was supported by the results, which showed that β -catenin Y333F behaved similarly to wild-type β -catenin both in binding to APC, AXIN2 and E-cadherin and in Wnt-induced β -catenin transactivation and cellular functions (Supplementary Fig. 10a–g). In contrast, expression of β -catenin Y333F blocked EGF- but not WNT3A-induced cell migration (Supplementary Fig. 10h).

To investigate the mechanisms underlying PKM2- and β -catenin-dependent cyclin D1 expression, we performed ChIP analyses. Figure 3e shows that EGF induced an enhanced binding of wild-type Flag-PKM2 and Flag-PKM2 K367M, but not of Flag-PKM2 K433E, to the *CCND1* promoter region. In addition, binding of Flag-PKM2 to the promoter region was not detected in ChIP analyses after β -catenin had been immunodepleted from cell lysates (Fig. 3f). These results indicate that the PKM2- β -catenin interaction, but not PKM2 kinase activity, is required for both proteins to bind to the *CCND1* promoter region.

We next examined whether PKM2 binding to the *CCND1* promoter region regulates histone H3 acetylation, which is important for gene transcription. ChIP analyses showed that EGF treatment resulted in a significant increase of histone H3 acetylation in the *CCND1* promoter region, which was blocked by PKM2 depletion (Fig. 3g). Reconstituted expression of rPKM2 K433E and the inactive rPKM2 K367M mutant failed to restore EGF-induced histone H3 acetylation, compared with the wild-type protein.

To understand further the mechanism underlying PKM2-regulated histone H3 acetylation, we performed ChIP analyses with antibodies against ubiquitously expressed histone deacetylase (HDAC)1, HDAC2 and HDAC3 (ref. 14). Figure 3h shows that HDAC3, but not HDAC1 or

HDAC2 (Supplementary Fig. 11a), was pre-bound to the *CCND1* promoter. EGF treatment resulted in the disassociation of HDAC3 from the *CCND1* promoter, and this disassociation was blocked by PKM2 depletion. Furthermore, reconstituted expression of wild-type rPKM2, but not of rPKM2 K367M or rPKM2 K433E mutants, was able to restore EGF-induced disassociation of HDAC3 from the promoter (Fig. 3h). These results indicate that the kinase activity of PKM2 and its binding with β -catenin to the *CCND1* promoter region are required for HDAC3 removal from the promoter. An additional co-immunoprecipitation analysis showed that wild-type PKM2, but not PKM2 K367M, interacted with HDAC3 upon EGF treatment (Supplementary Fig. 11b). These results suggest that PKM2 can interact with HDAC3 in an active conformation, which facilitates HDAC3 removal from the *CCND1* promoter, acetylation of histone H3, and cyclin D1 expression.

To support the findings that EGF-induced and c-Src-dependent PKM2- β -catenin interaction and subsequent cyclin D1 expression are not cell-line specific, we treated GSC11 and GSC23 human primary GBM cells with EGF. Supplementary Fig. 12a shows that EGF treatment results in nuclear translocation of PKM2 and cyclin D1 expression in these cells. In addition, EGF-induced phosphorylation of β -catenin Y333 (Supplementary Fig. 12b) and its association with PKM2 (Supplementary Fig. 12c) were blocked by pre-treatment with SU6656.

We next examined the significance of the PKM2- β -catenin interaction in tumour cell proliferation. Figure 4a shows that U87/EGFRvIII cells proliferated much faster than did parental U87 cells. Expression of β -catenin shRNA (Fig. 3c) largely inhibited U87/EGFRvIII cell proliferation, which was rescued by reconstituted expression of wild-type r β -catenin, but not of the r β -catenin Y333F mutant (Fig. 4a). In addition, EGFRvIII-promoted cell proliferation was similarly inhibited by PKM2 depletion, which was rescued by the reconstituted expression of wild-type rPKM2, but not of the rPKM2 K433E or rPKM2 K367M mutant (Supplementary Fig. 13a and Fig. 4b).

β -Catenin-regulated cyclin D1 expression is critical for G1-S phase transition and cell cycle progression¹⁵. Depletion of β -catenin or PKM2 resulted in an accumulation of U87/EGFRvIII cells in G0/G1 phase, which was rescued by reconstituted expression of wild-type r β -catenin or wild-type rPKM2, but not of the r β -catenin Y333F or rPKM2 K433E mutant (Supplementary Fig. 13b). Thus, the PKM2- β -catenin interaction is essential for cell cycle progression.

To determine the role of PKM2-dependent β -catenin transactivation in brain tumour development, we intracranially injected U87 or U87/EGFRvIII cells into athymic nude mice. U87 cells did not form a

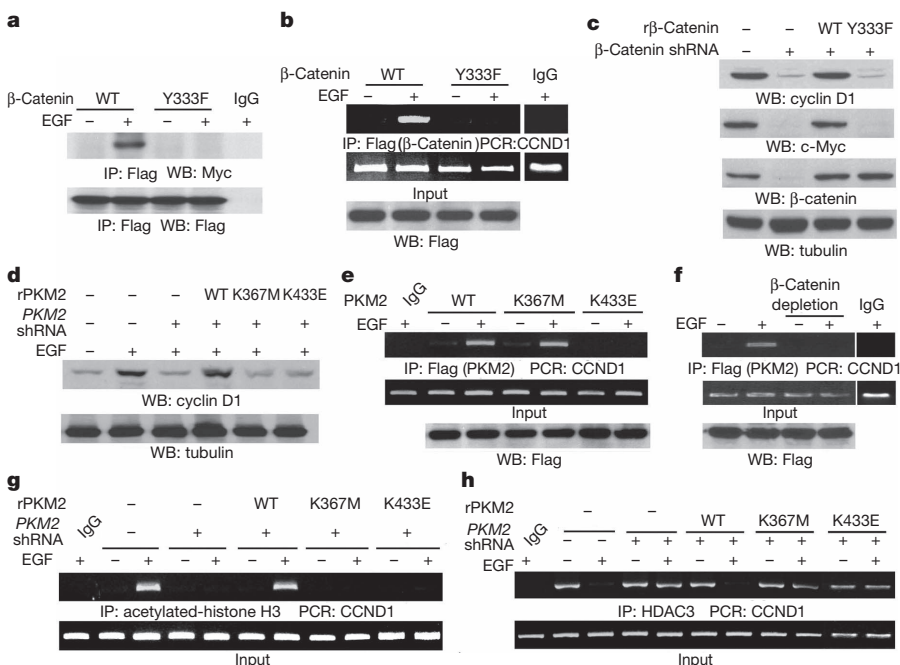


Figure 3 | The PKM2- β -catenin interaction is required for β -catenin-induced cyclin D1 expression. **a, b**, U87/EGFR cells transiently expressing Flag- β -catenin proteins were treated with or without EGF for 10 h. **c**, β -Catenin was depleted in U87/EGFRvIII cells, followed by reconstituted expression of r β -catenin. **d, g, h**, U87/EGFR cells with or without depleted PKM2 and reconstituted expression of rPKM2 were treated with or without EGF for 24 h (**d**) or 10 h (**g, h**). **e**, U87/EGFR cells transiently expressing the indicated Flag-tagged PKM2 protein were treated with or without EGF for 10 h. **f**, U87/EGFR cells transiently expressing Flag-PKM2 were treated with or without EGF for 10 h. β -Catenin was immunoprecipitated from the cell lysates, and the remaining supernatant was used for ChIP analyses.

detectable tumour 2 weeks after injection (Fig. 4c, bottom left). In contrast, U87/EGFRvIII cells elicited rapid tumorigenesis (Fig. 4c, top left). Notably, depletion of β -catenin (Fig. 4c, top) or PKM2 (Fig. 4c, bottom) abrogated EGFRvIII-driven tumour growth, which was rescued by expression of wild-type β -catenin or wild-type rPKM2, but not of the β -catenin Y333F or rPKM2 K433E mutant. Similar results were obtained using GSC11 cells (Supplementary Fig. 14a). In addition, adding doxycycline to the drinking water 14 days after intracranial injection of GSC11 cells that expressed a tetracycline-inducible *PKM2* shRNA partially depleted PKM2 expression in tumour tissues and inhibited tumour growth (Supplementary Fig. 14b). Thus, the PKM2– β -catenin interaction and PKM2 expression is instrumental in tumour growth.

To support further the role of c-Src in EGFR-induced β -catenin transactivation *in vivo*, we injected SU6656 intratumorally, which significantly blocked tumour growth (Supplementary Fig. 15a, b), reduced the phosphorylation levels of c-Src Y418 and β -catenin Y333, and inhibited cyclin D1 expression in tumour tissue (Supplementary Fig. 15c). The requirement of β -catenin transactivation in tumorigenesis was also examined by expression of WNT1 in U87/EGFRvIII-PKM2 shRNA cells. WNT1 expression resulted in the induction of cyclin D1 (Supplementary Fig. 16a) and largely rescued

PKM2 depletion-blocked tumorigenesis (Supplementary Fig. 16b, c). In addition, this effect was further enhanced using WNT1-expressing U87/EGFRvIII cells with reconstituted expression of rPKM2 K433E, which retains its catalytic activity for glycolysis. These results suggest that, whereas the metabolic function of PKM2 has a critical role in aerobic glycolysis⁷ and tumorigenesis, brain tumour development promoted by EGFR requires PKM2-modulated β -catenin transactivation.

Analysis of publicly available microarray data sets (Affymetrix, U133) from The Cancer Genome Atlas (TCGA) and other sources^{16–19} revealed a correlation of *c-Myc* and *CCND1* expression with *EGFR* expression in GBM samples (Supplementary Fig. 17). In addition, phosphorylation levels of β -catenin Y333 correlated with phosphorylation levels of activated c-Src in seven human primary GBM cell lines (Supplementary Fig. 18).

We next performed immunohistochemical (IHC) analyses to examine c-Src activity, β -catenin Y333 phosphorylation, and PKM2 nuclear localization in serial sections of 55 human primary GBM specimens by using antibodies with validated specificities (Supplementary Fig. 19). Figure 4d shows that the levels of c-Src Y418 phosphorylation, β -catenin Y333 phosphorylation, and nuclear PKM2 expression were correlated with each other. In addition, β -catenin Y333 phosphorylation accumulated in the nuclei of a large percentage

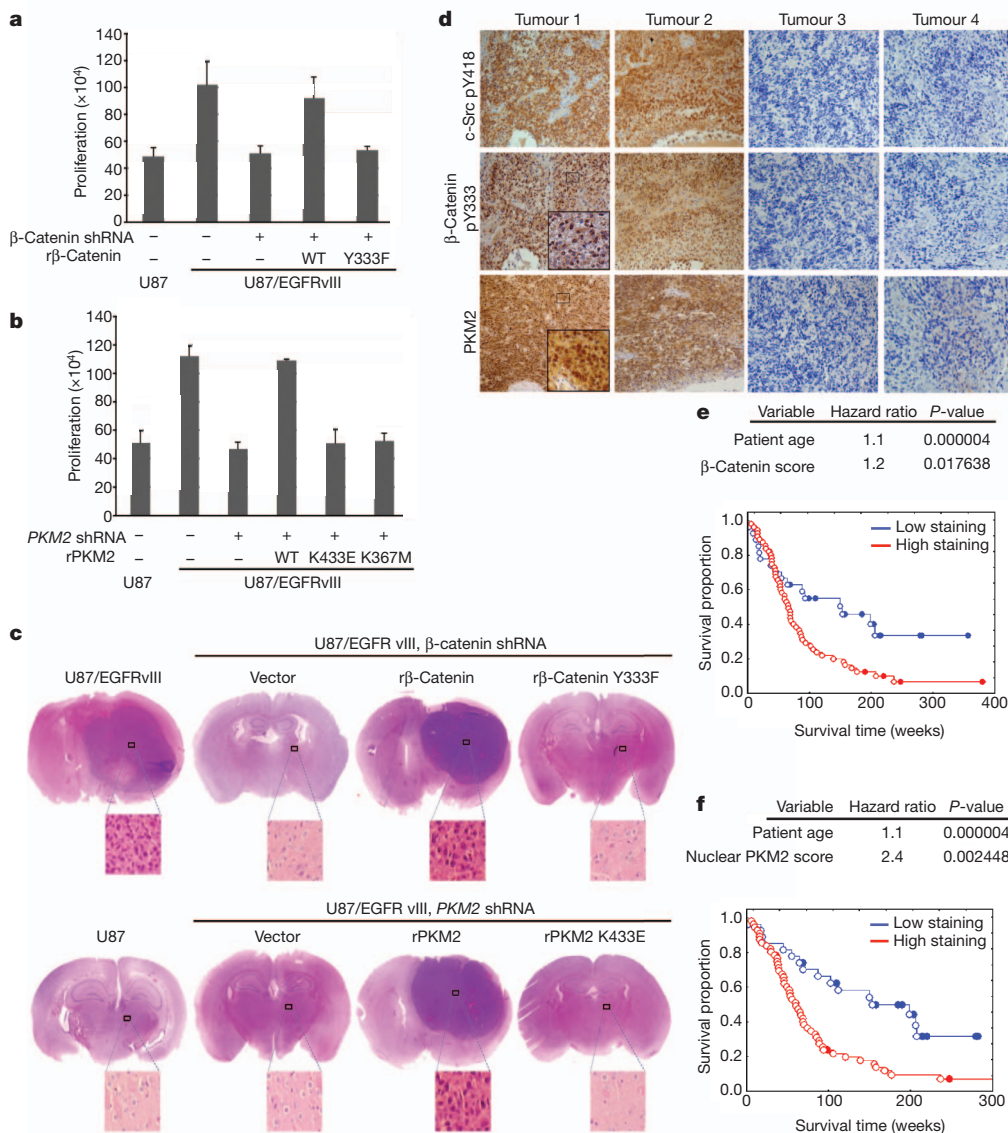


Figure 4 | The PKM2– β -catenin interaction is required for tumour development. **a**, **b**, U87, U87/EGFRvIII cells with or without depleted β -catenin and reconstituted expression of r β -catenin (**a**), or U87/EGFRvIII cells with or without depleted PKM2 and reconstitution of the expression of rPKM2 (**b**), were plated and counted 7 days after seeding. Data represent the means \pm s.d. of three independent experiments. **c**, U87 (bottom left), U87/EGFRvIII cells with or without depleted β -catenin and reconstituted expression of r β -catenin (top), or U87/EGFRvIII cells with or without depleted PKM2 and reconstituted expression of rPKM2 (bottom right) were intracranially injected into athymic nude mice. After 2 weeks, tumour growth was examined. H&E-stained coronal brain sections show representative tumour xenografts. **d**, IHC staining with the indicated antibodies was performed on 55 GBM specimens. Representative photos of four tumours are shown. **e**, **f**, The survival time for 84 patients with low (0–5 staining scores, blue curve) versus high (6–8 staining scores, red curve) β -catenin Y333 phosphorylation (**e**; low, 28 patients; high, 56 patients) and nuclear PKM2 expression (**f**; low, 28 patients; high, 56 patients) were compared (bottom). The tables (top) show the multivariate analysis after adjustment for patient age, indicating the significance level of the association of Y333-phosphorylated β -catenin expression (**e**) or nuclear PKM2 expression (**f**) with patient survival. Empty circles represent the deceased patients, and filled circles represent the censored (alive at last clinical follow-up) patients.

of tumour cells (Fig. 4d, left). Quantification of the staining upon a scale of 0 to 8.0 showed that these correlations were significant (Supplementary Fig. 20). The survival durations of 84 patients, all of whom received standard adjuvant radiotherapy after surgery, followed by treatment with an alkylating agent (temozolomide in the majority of cases), with low (0–5 staining) versus high (5.1–8 staining) β -catenin Y333 phosphorylation and nuclear PKM2 expression were compared. Patients whose tumours had low β -catenin Y333 phosphorylation or nuclear PKM2 expression had a median survival of 185.2 and 130.0 weeks, respectively. The median survival of patients decreased to 69.4 and 82.5 weeks, respectively, when their tumours showed high levels of β -catenin Y333 phosphorylation or nuclear PKM2 expression. In a Cox multivariate model, the IHC scores of β -catenin phosphorylation (Fig. 4e) and nuclear PKM2 expression (Fig. 4f) were independent predictors of glioblastoma patient survival, after adjusting for the age of the patient, a relevant clinical covariate. These results support the role of PKM2 association with c-Src-phosphorylated β -catenin Y333 in the clinical behaviour of human GBM and reveal a relationship between β -catenin Y333 phosphorylation/nuclear PKM2 localization and clinical aggressiveness of the tumour. These findings were further supported by IHC analyses, showing significantly lower levels of β -catenin Y333 phosphorylation in low-grade diffuse astrocytoma (30 cases) (World Health Organization (WHO) grade II; median survival time >5 years) than were present in GBM specimens (WHO grade IV)²⁰ (Supplementary Fig. 21).

We previously showed that GSK-3 β -independent transactivation of β -catenin by growth factor receptor occurs by mechanisms distinct from Wnt-dependent canonical signalling^{4,21–23}. In this study, we describe an important and previously unknown mechanism underlying EGFR activation-induced β -catenin transactivation through interaction with PKM2, which has a critical role in transcription of *CCND1* and *c-Myc* (Supplementary Fig. 1). The understanding that phosphorylation of β -catenin Y333 and its interaction with PKM2 are required for tumour cell proliferation and tumour development, and that the levels of β -catenin Y333 phosphorylation and nuclear PKM2 correlate with grades of glioma malignancy and prognosis, may provide a molecular basis for improved diagnosis and treatment of tumours with activated EGFR and upregulated PKM2. Our findings, in combination with previous reports^{7,24}, delineate two essential mechanisms underlying tumour development by regulation of metabolic and non-metabolic functions of PKM2: (1) PKM2 enhances aerobic glycolysis^{7,24}; and (2) PKM2 promotes tumour cell proliferation by binding to and transactivating Y333-phosphorylated β -catenin. Thus, PKM2 has dual roles that are essential for tumorigenesis: regulating cancer cell metabolism and gene transcription required for cell proliferation. The coordinated control of metabolism and proliferation by PKM2 is essential for tumorigenesis.

METHODS SUMMARY

Detailed methodology can be found in Methods. In short, the luciferase reporter gene assay, immunoprecipitation and immunoblotting analysis, purification of recombinant proteins, *in vitro* kinase assays and cell migration assay were performed as previously described^{14,21,22}.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 10 January; accepted 27 September 2011.

Published online 6 November; corrected 1 December 2011 (see full-text HTML version for details).

1. Vander Heiden, M. G., Cantley, L. C. & Thompson, C. B. Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* **324**, 1029–1033 (2009).

2. Cairns, R. A., Harris, I. S. & Mak, T. W. Regulation of cancer cell metabolism. *Nature Rev. Cancer* **11**, 85–95 (2011).
3. Koppenol, W. H., Bounds, P. L. & Dang, C. V. Otto Warburg's contributions to current concepts of cancer metabolism. *Nature Rev. Cancer* **11**, 325–337 (2011).
4. Lu, Z. & Hunter, T. Wnt-independent β -catenin transactivation in tumor development. *Cell Cycle* **3**, 569–571 (2004).
5. Lu, Z., Jiang, G., Blume-Jensen, P. & Hunter, T. Epidermal growth factor-induced tumor cell invasion and metastasis initiated by dephosphorylation and downregulation of focal adhesion kinase. *Mol. Cell. Biol.* **21**, 4016–4031 (2001).
6. Wykosky, J., Fenton, T., Furnari, F. & Cavenee, W. K. Therapeutic targeting of epidermal growth factor receptor in human cancer: successes and limitations. *Chin. J. Cancer* **30**, 5–12 (2011).
7. Christofk, H. R. et al. The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. *Nature* **452**, 230–233 (2008).
8. Yochum, G. S. et al. Serial analysis of chromatin occupancy identifies β -catenin target genes in colorectal carcinoma cells. *Proc. Natl Acad. Sci. USA* **104**, 3324–3329 (2007).
9. Christofk, H. R., Vander Heiden, M. G., Wu, N., Asara, J. M. & Cantley, L. C. Pyruvate kinase M2 is a phosphotyrosine-binding protein. *Nature* **452**, 181–186 (2008).
10. Le Mellay, V. et al. Regulation of glycolysis by Raf protein serine/threonine kinases. *Adv. Enzyme Regul.* **42**, 317–332 (2002).
11. Mazurek, S. Pyruvate kinase type M2: a key regulator of the metabolic budget system in tumor cells. *Int. J. Biochem. Cell Biol.* **43**, 969–980 (2011).
12. Coluccia, A. M. et al. Bcr-Abl stabilizes β -catenin in chronic myeloid leukemia through its tyrosine phosphorylation. *EMBO J.* **26**, 1456–1466 (2007).
13. Miravet, S. et al. Tyrosine phosphorylation of plakoglobin causes contrary effects on its association with desmosomes and adherens junction components and modulates β -catenin-mediated transcription. *Mol. Cell. Biol.* **23**, 7391–7402 (2003).
14. Xia, Y. et al. c-Jun downregulation by HDAC3-dependent transcriptional repression promotes osmotic stress-induced cell apoptosis. *Mol. Cell* **25**, 219–232 (2007).
15. Tetsu, O. & McCormick, F. β -Catenin regulates expression of cyclin D1 in colon carcinoma cells. *Nature* **398**, 422–426 (1999).
16. Freije, W. A. et al. Gene expression profiling of gliomas strongly predicts survival. *Cancer Res.* **64**, 6503–6510 (2004).
17. Gravendeel, L. A. et al. Intrinsic gene expression profiles of gliomas are a better predictor of survival than histology. *Cancer Res.* **69**, 9065–9072 (2009).
18. Petalidis, L. P. et al. Improved grading and survival prediction of human astrocytic brain tumors by artificial neural network analysis of gene expression microarray data. *Mol. Cancer Ther.* **7**, 1013–1024 (2008).
19. Phillips, H. S. et al. Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis. *Cancer Cell* **9**, 157–173 (2006).
20. Furnari, F. B. et al. Malignant astrocytic glioma: genetics, biology, and paths to treatment. *Genes Dev.* **21**, 2683–2710 (2007).
21. Ji, H. et al. EGF-induced ERK activation promotes CK2-mediated disassociation of α -catenin from β -catenin and transactivation of β -catenin. *Mol. Cell* **36**, 547–559 (2009).
22. Fang, D. et al. Phosphorylation of β -catenin by AKT promotes β -catenin transcriptional activity. *J. Biol. Chem.* **282**, 11221–11229 (2007).
23. Lu, Z., Ghosh, S., Wang, Z. & Hunter, T. Downregulation of caveolin-1 function by EGF leads to the loss of E-cadherin, increased transcriptional activity of β -catenin, and enhanced tumor cell invasion. *Cancer Cell* **4**, 499–515 (2003).
24. Luo, W. et al. Pyruvate kinase M2 is a PHD3-stimulated coactivator for hypoxia-inducible factor 1. *Cell* **145**, 732–744 (2011).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank T. Hunter (The Salk Institute for Biological Studies) for Abl and Src knockout cells, H. Clevers (Netherlands Institute for Developmental Biology) for the pTOP-FLASH and the pFOP-FLASH, and Y. Li (Baylor College of Medicine) for a WNT1 lenti-vector. This work was supported by National Cancer Institute grants 5R01CA109035 (Z.L.), 5 P50 CA127001-03 and CA16672 (Cancer Center Support Grant); a research grant (RP110252; Z.L.) from the Cancer Prevention and Research Institute of Texas (CPRIT), an American Cancer Society Research Scholar Award RSG-09-277-01-CSM (Z.L.), and a Sister Institution Network Fund from The University of Texas MD Anderson Cancer Center (Z.L.).

Author Contributions This study was conceived by Z.L. Z.L. and W.Y. designed the study; W.Y., Y.X., H.J., Y.Z. and J.L. performed experiments; K.A. provided pathology assistance; W.H. and X.G. provided reagents and conceptual advice; Z.L. wrote the paper with comments from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to Z.L. (zhiminlu@mdanderson.org).

METHODS

Cells and cell culture conditions. U87, U87/EGFR and U251 GBM cells; DU145 prostate cancer cells; MDA-MB-231 breast cancer cells; A431 epidermoid carcinoma cells and NIH3T3, 293T, c-SRC^{+/+}, c-SRC^{-/-}, ABL^{+/+} and ABL^{-/-} cells were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% bovine calf serum (HyClone). Human primary GBM cells were maintained in DMEM/F-12 50/50 supplemented with B27, EGF (10 ng ml⁻¹), bFGF (10 ng ml⁻¹). Rat pheochromocytoma PC12 cells were maintained in DMEM supplemented with 10% horse serum and 5% fetal bovine serum. Cell cultures were made quiescent by growing them to confluence, and the medium was replaced with fresh medium containing 0.5% serum for 1 day. EGF with 100 ng ml⁻¹ final concentration was used for cell stimulation.

Materials. Rabbit polyclonal antibodies recognizing phospho- β -catenin Y333, PKM1, PKM2, c-Src, Abl, phospho-Abl Y412 and c-Myc were obtained from Signallway Biotechnology. A mouse antibody recognizing phospho-tyrosine was obtained from BD Biosciences. A monoclonal antibody against phospho-c-Src Y418 was purchased from Millipore. Polyclonal antibodies for cyclin D1 and PCNA and monoclonal antibodies for β -catenin and Myc tag were purchased from Santa Cruz Biotechnology. A polyclonal antibody of acetylated histone H3 and a monoclonal antibody for HDAC3 were obtained from Upstate Biotechnology. EGF and mouse monoclonal antibodies for Flag, GST, His and tubulin were purchased from Sigma. Hygromycin, puromycin, G418, SU6656, Abl inhibitor, cycloheximide, DNase-free RNase A and propidium iodide were purchased from EMD Biosciences. Active c-Src was obtained from Signallchem. Hoechst 33342 and Alexa Fluor 488 goat anti-rabbit antibody was from Molecular Probes. HyFect transfection reagents were from Denville Scientific. GelCode Blue Stain Reagent was obtained from Pierce.

Transfection. Cells were plated at a density of 4×10^5 per 60-mm dish 18 h before transfection. Transfection was performed as previously described¹⁴.

Immunoprecipitation and immunoblotting analysis. Extraction of proteins with a modified buffer from cultured cells was followed by immunoprecipitation and immunoblotting with corresponding antibodies, as described previously²⁵.

Cell proliferation assay. 2×10^4 cells were plated and counted 7 days after seeding in DMEM with 0.5% bovine calf serum. Data represent the means \pm s.d. of three independent experiments.

DNA constructs and mutagenesis. PCR-amplified human PKM1 was cloned into pcDNA3.1/hygro (+) vector between BamHI and XhoI. PCR-amplified human PKM2 was cloned into pcDNA3.1/hygro (+) vector between BamHI and NotI. β -Catenin was subcloned into pGEX-4T-1 vector between BamHI and NotI. pcDNA 3.1/hygro (+)-PKM2, -K433E, -K367M, pcDNA 3.1/hygro (+)- β -catenin Y86F, and - β -catenin Y333F were made using the QuikChange site-directed mutagenesis kit (Stratagene). pcDNA 3.1-rPKM2 contains mutations of C1192T and C1194G.

The pGIPZ control was generated with a control oligonucleotide GCTTCTAACACCGGAGGTCTT. pGIPZ PKM2 shRNA and pTRIPZ PKM2 shRNA were generated with TTATTTGAGGAAGTCCGCCG oligonucleotide targeting exon 10 of the PKM2 transcript. pGIPZ β -catenin shRNA was generated with CATGCACAAGAATGGATCACAA.

Flow cytometry analysis. 1×10^6 treated cells were fixed in cold 70% ethanol for 3 h, spun down, and incubated for 1 h at 37 °C in PBS with DNase-free RNase A (100 μ g ml⁻¹) and propidium iodide (50 μ g ml⁻¹). Cells were then analysed by fluorescence-activated cell sorting (FACS).

Purification of recombinant proteins. The wild-type and mutants of GST- β -catenin were expressed in bacteria and purified as described previously¹⁴.

In vitro kinase assays. The kinase reactions were performed as described previously²².

Luciferase reporter gene assay. The transcriptional activation of β -catenin in 293T cells was measured as previously described²². The relative levels of luciferase activity were normalized to the levels of untreated cells and to the levels of luciferase activity of the *Renilla* control plasmid.

ChIP assay. ChIP was performed using an Upstate Biotechnology kit. Chromatin prepared from cells (in a 10-cm dish) was used to determine total DNA input and for overnight incubation with the specific antibodies or with normal rabbit or mouse immunoglobulin G. The human *CCND1* promoter-specific primers used in PCR were 5'-GGGGCGATTTCATTCTAT-3' (forward) and 5'-CGGTCTGTTGAGGAGTTGG-3' (reverse).

Immunofluorescence analysis. Immunofluorescence analysis was performed as described previously²².

Subcellular fractionation. Nuclei, cytosol and cell membranes were isolated using the Nuclear Extract Kit from Active Motif North America and the ProteoExtract Subcellular Proteome Extraction Kit from Calbiochem.

Immunohistochemical analysis. Mouse tumour tissues were fixed and prepared for staining. The specimens were stained with Mayer's haematoxylin

and subsequently with eosin (Biogenex Laboratories). Afterwards, the slides were mounted using Universal Mount (Research Genetics).

The tissue sections from paraffin-embedded human GBM specimens were stained with antibodies against phospho-c-Src Y418, phospho- β -catenin Y333, PKM2, or nonspecific IgG as a negative control. We quantitatively scored the tissue sections according to the percentage of positive cells and staining intensity, as previously defined²¹. We assigned the following proportion scores: 0 if 0% of the tumour cells showed positive staining, 1 if 0% to 1% of cells were stained, 2 if 1% to 10% stained, 3 if 11% to 30% stained, 4 if 31% to 70% stained, and 5 if 71% to 100% stained. We rated the intensity of staining on a scale of 0 to 3: 0, negative; 1, weak; 2, moderate; and 3, strong. We then combined the proportion and intensity scores to obtain a total score (range, 0–8), as described previously²¹. Scores were compared with overall survival, defined as the time from date of diagnosis to death or last known date of follow-up. All patients received standard adjuvant radiotherapy after surgery, followed by treatment with an alkylating agent (temozolomide in the majority of cases). The use of human brain tumour specimens and the database was approved by the institutional review board at MD Anderson Cancer Center.

Intracranial injection. We intracranially injected 5×10^5 GBM cells (in 5 μ l of DMEM per mouse), with or without regulation of β -catenin or PKM2 expression, into 4-week-old female athymic nude mice. The intracranial injections were performed as described in a previous publication²⁶. Seven mice per group in each experiment were included. Animals injected with U87/EGFRvIII or GSC 11 cells were killed 2 weeks or 30 days after glioma cell injection, respectively. The brain of each mouse was harvested, fixed in 4% formaldehyde, and embedded in paraffin. Tumour formation and phenotype were determined by histological analysis of H&E-stained sections.

For doxycycline induction studies, 7 mice were killed 14 days after GBM cell injection to examine tumour growth, whereas the remaining 14 mice in two groups were fed with or without water containing 800 μ g ml⁻¹ doxycycline. The water containing doxycycline was changed every 3 days. The mice in these two groups were killed at day 30. For the SU6656 treatment studies, 7 mice were killed 5 days after U87/EGFRvIII cell injection to examine tumour growth, whereas the remaining 14 mice in two groups were treated with either DMSO or SU6656. SU6656 (0.015 mg kg⁻¹ in 5 μ l of DMSO) was intracranially injected into the tumour every 3 days, and the mice in both groups were killed at day 14 after GBM cell injection.

Neurite extension assay. Mouse NGF (Upstate Biotechnology) was added at 100 ng ml⁻¹ to PC12 cells in culture, and the medium was changed every 3 days. Two hundred cells from 10 randomly chosen microscopic fields for each condition were examined. A cell was considered to respond to NGF if it extended neurites at least two cell bodies long after the incubation duration.

Pyruvate kinase assay. The activity of bacterially purified wild-type PKM2 (0.1 μ g) and PKM2 K433E (0.1 μ g) towards PEP was measured by a pyruvate kinase assay (BioVision), according to the manufacturer's instruction. Data represent the mean \pm s.d. of three independent experiments.

Cell migration assay. Matrigel-transwell assay was performed as described previously²².

Lentivirus preparation. We transfected 293T cells in 150-mm plates with 12 μ g pFU-CGW WNT1, 4 μ g pMDL, 4 μ g pRSV-Rev, and 4 μ g pCMV-VSVG, a plasmid encoding the G-protein of the vesicular stomatitis virus (VSV-G) envelope. The medium was changed the next day. The medium containing lentivirus was harvested at 48 h and 72 h after transfection. Virus particles were concentrated and purified by ultra-high-speed centrifugation (25,000g for 2 h at 4 °C). Cells were infected with lentivirus (1×10^6) in the presence of 6 μ g ml⁻¹ polybrene (Sigma).

Quantitative real-time PCR. Total RNA was extracted using a RNA High-purity Total RNA Rapid Extraction Kit (Signallway Biotechnology). cDNA was prepared using oligonucleotide (dT), random primers, and a Thermo Reverse Transcription kit (Signallway Biotechnology). Quantitative real-time PCR analysis was performed using 2 \times SIBR real-time PCR Premixture (Signallway Biotechnology) under the following conditions: 5 min at 95 °C followed by 40 cycles at 95 °C for 30 s, 55 °C for 40 s, and 72 °C for 1 min using an ABI Prism 7700 sequence detection system. Data were normalized to expression of a control gene (*b-actin*) for each experiment.

The following primer pairs were used for quantitative real-time PCR: DKK1, 5'-CATTGACAACCTACAGCCGTAC-3' (forward) and 5'-GGGCAGACATAGCGTGA-3' (reverse); AXIN2, 5'-GGGAGCCTAAAGTCTGTG-3' (forward) and 5'-GGGTTCTCGGAAATGAG-3' (reverse); β TRCP, 5'-CCCCAACTGACATTACCC-3' (forward) and 5'-TCGAATACAACGCACCAA-3' (reverse); β -actin, 5'-ATGGATGACGATATCGTCTGCGC-3' (forward) and 5'-GCAGCACAGGTGCTCTCA-3' (reverse).

25. Lu, Z. *et al.* Activation of protein kinase C triggers its ubiquitination and degradation. *Mol. Cell. Biol.* **18**, 839–845 (1998).

26. Gomez-Manzano, C. *et al.* Delta-24 increases the expression and activity of topoisomerase I and enhances the antitumour effect of irinotecan. *Clin. Cancer Res.* **12**, 556–562 (2006).

Control of *Drosophila* endocycles by E2F and CRL4^{CDT2}

Norman Zielke^{1,2*}, Kerry J. Kim^{3*}, Vuong Tran^{2*}, Shusaku T. Shibutani⁴, Maria-Jose Bravo², Sabarish Nagarajan^{5,6}, Monique van Straaten¹, Brigitte Woods², George von Dassow³, Carmen Rottig⁷, Christian F. Lehner⁷, Savraj S. Grewal^{5,6}, Robert J. Duronio⁴ & Bruce A. Edgar^{1,2}

Endocycles are variant cell cycles comprised of DNA synthesis (S-) and gap (G)-phases but lacking mitosis^{1,2}. Such cycles facilitate post-mitotic growth in many invertebrate and plant cells, and are so ubiquitous that they may account for up to half the world's biomass^{3,4}. DNA replication in endocycling *Drosophila* cells is triggered by cyclin E/cyclin dependent kinase 2 (CYCE/CDK2), but this kinase must be inactivated during each G-phase to allow the assembly of pre-Replication Complexes (preRCs) for the next S-phase^{5,6}. How CYCE/CDK2 is periodically silenced to allow re-replication has not been established. Here, using genetic tests in parallel with computational modelling, we show that the endocycles of *Drosophila* are driven by a molecular oscillator in which the E2F1 transcription factor promotes *CycE* expression and S-phase initiation, S-phase then activates the CRL4^{CDT2} ubiquitin ligase, and this in turn mediates the destruction of E2F1 (ref. 7). We propose that it is the transient loss of E2F1 during S phases that creates the window of low Cdk activity required for preRC formation. In support of this model overexpressed E2F1 accelerated endocycling, whereas a stabilized variant of E2F1 blocked endocycling by deregulating target genes, including *CycE*, as well as *Cdk1* and mitotic cyclins. Moreover, we find that altering cell growth by changing nutrition or target of rapamycin (TOR) signalling impacts E2F1 translation, thereby making endocycle progression growth-dependent. Many of the regulatory interactions essential to this novel cell cycle oscillator are conserved in animals and plants^{1,2,8}, indicating that elements of this mechanism act in most growth-dependent cell cycles.

S-phase control in proliferating animal cells depends on the E3 ubiquitin ligase, APC^{FZY} (FZY is also known as CDC20), which is activated by a cyclin-CDK1 complex during mitosis. APC^{FZY} promotes the degradation of mitotic cyclins, thereby extinguishing CDK1 activity following mitosis, and it also promotes the degradation of geminin (also known as Gem), an inhibitor of the preRC component Cdt1. The combination of low geminin and low cyclin-CDK1 activity during early G1 allows the assembly of preRCs containing origin recognition complex (ORC) proteins, CDC6, double parked (DUP), encodes the *Drosophila* orthologue of CDT1), and MCM2-7 onto replication origins, thus 'licensing' the DNA for renewed replication⁹. *Drosophila* endocycling cells do not express mitotic cyclin-CDK1 complexes or APC^{FZY} (refs 10, 11), and so this mechanism of S-phase regulation cannot apply to them. These endocycles do however use CYCE/CDK2 to trigger S-phases^{12,13} (Fig. 1, 2), and they also require the G1-specific APC variant, APC^{Fzr/Cdh1}, which mediates cyclic degradation of cell cycle factors including geminin and ORC1 (refs 10, 14). Importantly, whereas overexpressed CYCE/CDK2 is tolerated in

mitotic cell cycles¹⁵, it blocks endocycling (Fig. 2, 3)^{5,6}. This is likely due to CYCE/CDK2's ability to suppress APC^{Fzr/Cdh1} and drive geminin accumulation^{10,14,16}, although CYCE may also inhibit preRC formation directly by phosphorylating preRC components. The importance of CYCE oscillation for endocycling is underscored by the finding that

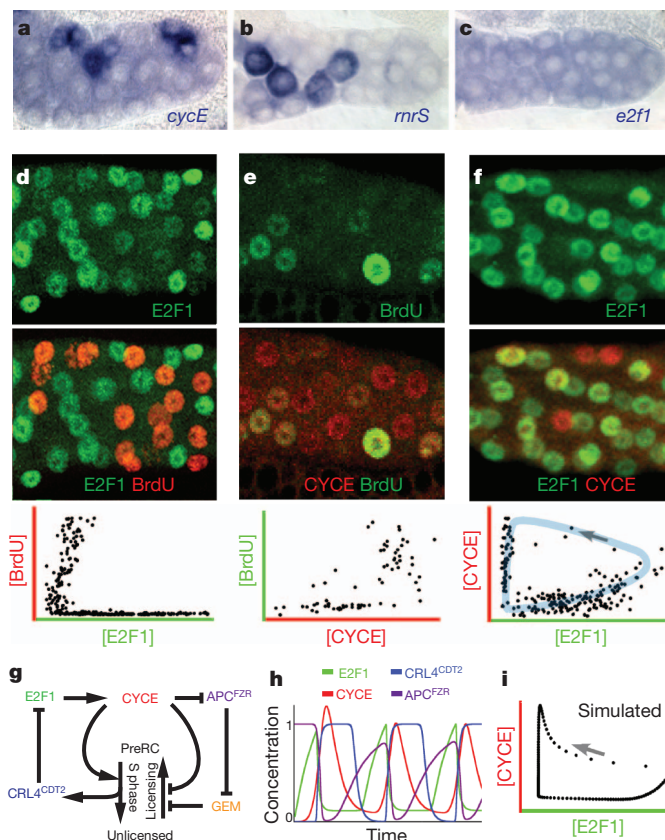


Figure 1 | Wild-type salivary gland endocycles. **a–c**, *In situ* hybridization of wild-type (WT) 72 h AED glands to the indicated mRNAs. **d–f**, WT salivary glands at 72 h AED double-labelled for E2F1 (green) and BrdU (red) (**d**), CYCE (red) and BrdU (green) (**e**), or CYCE (red) and E2F1 (green) (**f**). Graphs show nuclear concentrations measured from micrographs of two to three glands, in which each dot represents one nucleus. Shaded region (blue) shows trajectory of E2F1/CYCE oscillations with an arrow indicating the expected temporal progression. **g**, Simplified schematic of the computational model. See Supplementary Fig. 4. **h**, Time plot for WT predicted by the model. **i**, Nuclear concentrations predicted by the model; arrow represents temporal progression.

¹German Cancer Research Center (DKFZ)-Zentrum für Molekulare Biologie der Universität Heidelberg (ZMBH) Alliance, Im Neuenheimer Feld 282, 69120 Heidelberg, Germany. ²Fred Hutchinson Cancer Research Center, 1100 Fairview Avenue North, Seattle, Washington 98109, USA. ³Center for Cell Dynamics, Friday Harbor Labs, University of Washington, 620 University Road, Friday Harbor, Washington 98250, USA. ⁴Department of Biology, University of North Carolina, Chapel Hill, North Carolina 27599, USA. ⁵Clark H. Smith Brain Tumor Center, Southern Alberta Cancer Research Institute, University of Calgary, 3330 Hospital Drive, Calgary, Alberta T2N 4N1, Canada. ⁶Department of Biochemistry and Molecular Biology, University of Calgary, 3330 Hospital Drive, Calgary, Alberta T2N 4N1, Canada. ⁷Institute of Molecular and Life Sciences (IMLS), Universität Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland.

*These authors contributed equally to this work.

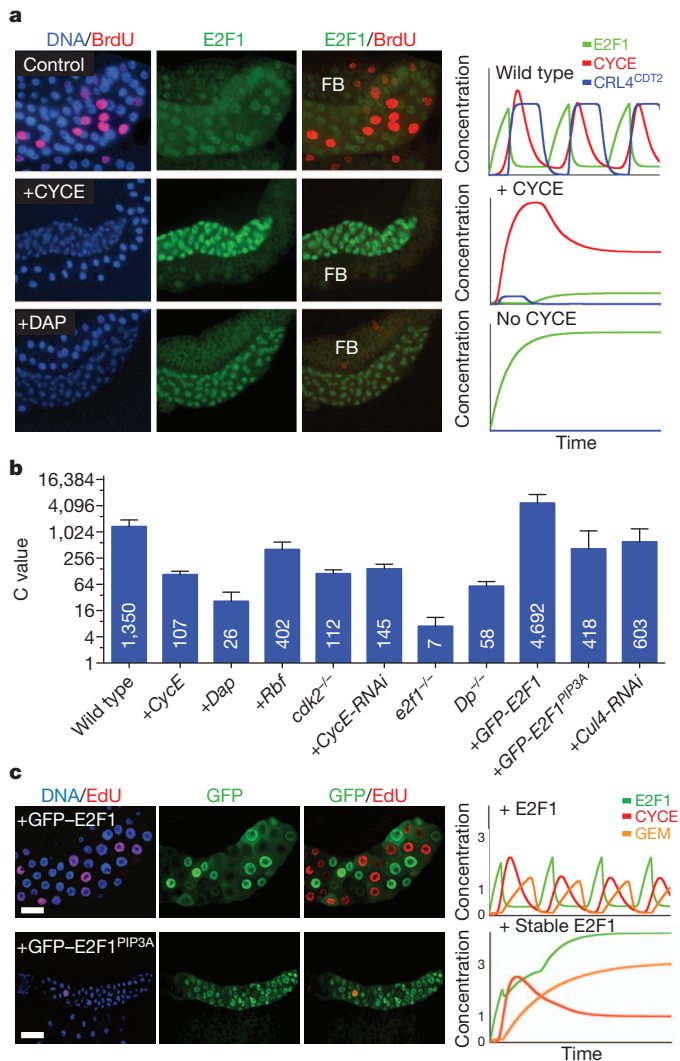


Figure 2 | Genetic tests of the endocycle mechanism. **a**, Salivary glands (centred) and associated fat body (above or below; FB) from 72 h AED larvae expressing the indicated genes under *ptc-Gal4/UAS* control. *ptc-Gal4* expresses in salivary glands but not in fat body. Left column shows DNA (blue) and BrdU (red) incorporated from 71–72 h AED. Middle column shows E2F1 (green). Right column shows E2F1 and BrdU. All images had identical exposures and magnifications. Graphs (right) show simulated time plots of E2F1 (green) CYCE (red) and CRL4^{CDT2} activity (blue) for each genotype. See Supplementary Table 1 for parameters. **b**, Nuclear DNA values from 96 h AED glands. For each genotype about 40 nuclei from 6–20 salivary glands were analysed. Error bars represent standard deviations. *ptc-Gal4* drove expression of the UAS-linked transgenes indicated with '+'. *Dp*^{-/-}, *Dp*^{g2/Df(2R)Exel7124 mutant. *E2f1*^{-/-}, *E2f1*⁷¹⁷² mutant cells generated by mitotic recombination. *cdk2*^{-/-} mutant glands were generated as described in methods. **c**, Salivary glands expressing wild-type GFP-E2F1 (above) or GFP-E2F1^{PIP3A} (below). Layout as in **a**. EdU, 5-ethynyl-2'-deoxyuridine.}

archipelago (AGO, also known as Cdc4 and Fbw7), which promotes CYCE degradation as a component of an SCF ubiquitin ligase, is required for the progression of endocycles, but not mitotic cycles (Supplementary Fig. 1)¹⁷. Despite its importance, the mechanism controlling CYCE/CDK2 periodicity in the endocycle has remained obscure for over a decade.

We addressed this problem in *Drosophila* larval salivary glands, which undergo ~10 asynchronous endocycles from ~7–96 h after egg deposition (h AED), reaching a final ploidy of ~1,350 C (ref. 18). Studies in the fly ovary had suggested that the CYCE/CDK2 inhibitor dacapo (DAP) might periodically silence CDK2 during endocycling¹⁹, but our analysis ruled this out for salivary glands (Supplementary Fig. 2)¹⁰. Hence we

asked whether cyclic CYCE/CDK2 activity might be controlled transcriptionally. *CycE* transcription is regulated by the E2F1 transcription factor^{20–22}, the accumulation of which is periodic in mitotic *Drosophila* cells^{7,23–25} because it is targeted for degradation during S phase by the PCNA/replication fork-associated E3 ubiquitin ligase CRL4^{CDT2} (ref. 7). In the salivary cells, *E2f1* mRNA was ubiquitous (Fig. 1c) but E2F1 protein was cyclic, being virtually absent in S-phase nuclei (Figs 1d, 2a). Continuously overexpressed E2F1 proteins were also depleted from S-phase nuclei (Figs 2c, 3c), consistent with periodic degradation. This implied that E2F-dependent transcription might also oscillate. Indeed, the mRNAs encoding *CycE* and two other E2F targets, *RnrS* and *pcna* (also known as *mus209*) were periodic (Figs 1, 3 and Supplementary Fig. 15, see also ref. 13). These mRNAs accumulated when E2F1 was overexpressed (Fig. 3a) and were reduced in mutants for *Dp*, the obligate dimerization partner of E2F1 (Supplementary Fig. 15). Thus periodic *CycE* expression is probably due to periodic activity of its regulator, E2F. CYCE protein was also cyclic, being present during a bit of each gap phase and much of each S-phase (Fig. 1e)^{13,26}. Based on these and other results¹⁰ we determined that E2F1 accumulates during G phases and is destroyed upon entry into S phase, whereas its target CYCE rises late in G phases and persists through most of each S phase.

These observations suggested that endocycles run using a molecular oscillator in which E2F1 promotes *CycE* transcription, and then CYCE/CDK2 triggers S-phase and the consequent destruction of E2F1 to reset the cycle (Fig. 1g). To evaluate this hypothesis we built a computational model that translated known regulatory interactions into a system of delay differential equations describing the concentrations of E2F1, RBF, CYCE, geminin and DUP, and the activities of APC^{Fzr/Cdh1} and CRL4^{CDT2} (Fig. 1g, Supplementary Methods and Supplementary Figs 4, 5). In this model, when CYCE was low geminin was degraded by APC^{Fzr/Cdh1}, allowing preRC licensing through DUP. High CYCE suppressed APC^{Fzr/Cdh1} activity and allowed geminin accumulation, and also triggered phosphorylation of RBF, S-phase initiation, activation of CRL4^{CDT2} and the subsequent degradation of E2F1 and DUP. The model's behaviour depended on unmeasured parameters representing biochemical kinetics (Supplementary Table 1), but Monte-Carlo searches found numerous parameters sets that simulated actual endocycles (Fig. 1h, i). The model robustly produced oscillations of its components despite quantitative parameter variation (Supplementary Figs 6–9) and did not require exquisitely tuned kinetics to reproduce oscillations like those observed *in vivo* (Supplementary Discussion).

We tested the computational model by challenging it to reproduce the results of genetic experiments performed in parallel. The model reproduced nearly all observed mutant and gene overexpression phenotypes (Fig. 2 and Supplementary Fig. 10). Notably, it predicted that increasing E2F1 should accelerate endocycling and lead to hyper-polyploidy, as subsequently observed experimentally (Figs 2, 3 and Supplementary Fig. 11). As predicted, we observed increased relative DNA amounts in *E2f1*^{+/+} cells generated in an *E2f1*^{+/−} background, and found that *E2f1*⁷¹⁷² homozygous null mutant cells supported essentially no endocycling (Fig. 2b and Supplementary Fig. 13). Thus both loss- and gain-of-function experiments indicated that E2F1 is an essential dose-dependent regulator of endocycle progression.

An important prediction of the computational model was that periodic E2F1 destruction should be essential for endocycling. *Drosophila* E2F1 is targeted for proteolysis during S phase via a conserved motif, the PIP box, which binds the replication fork-associated protein, PCNA, and mediates interaction with the CRL4^{CDT2} ubiquitin ligase⁷. Consistent with model predictions, a stabilized but active form of E2F1 lacking the PIP box (green fluorescent protein-conjugated GFP-E2F1^{PIP3A})⁷ blocked endocycle progression (Figs 2, 3 and Supplementary Fig. 11). Likewise RNA interference against CUL-4, a CRL4^{CDT2} component, arrested endocycling (Fig. 2b and Supplementary Figs 11 and 12). Levels of E2F1 in cells arrested by GFP-E2F1^{PIP3A} were not higher than in control GFP-E2F1-expressing cells that

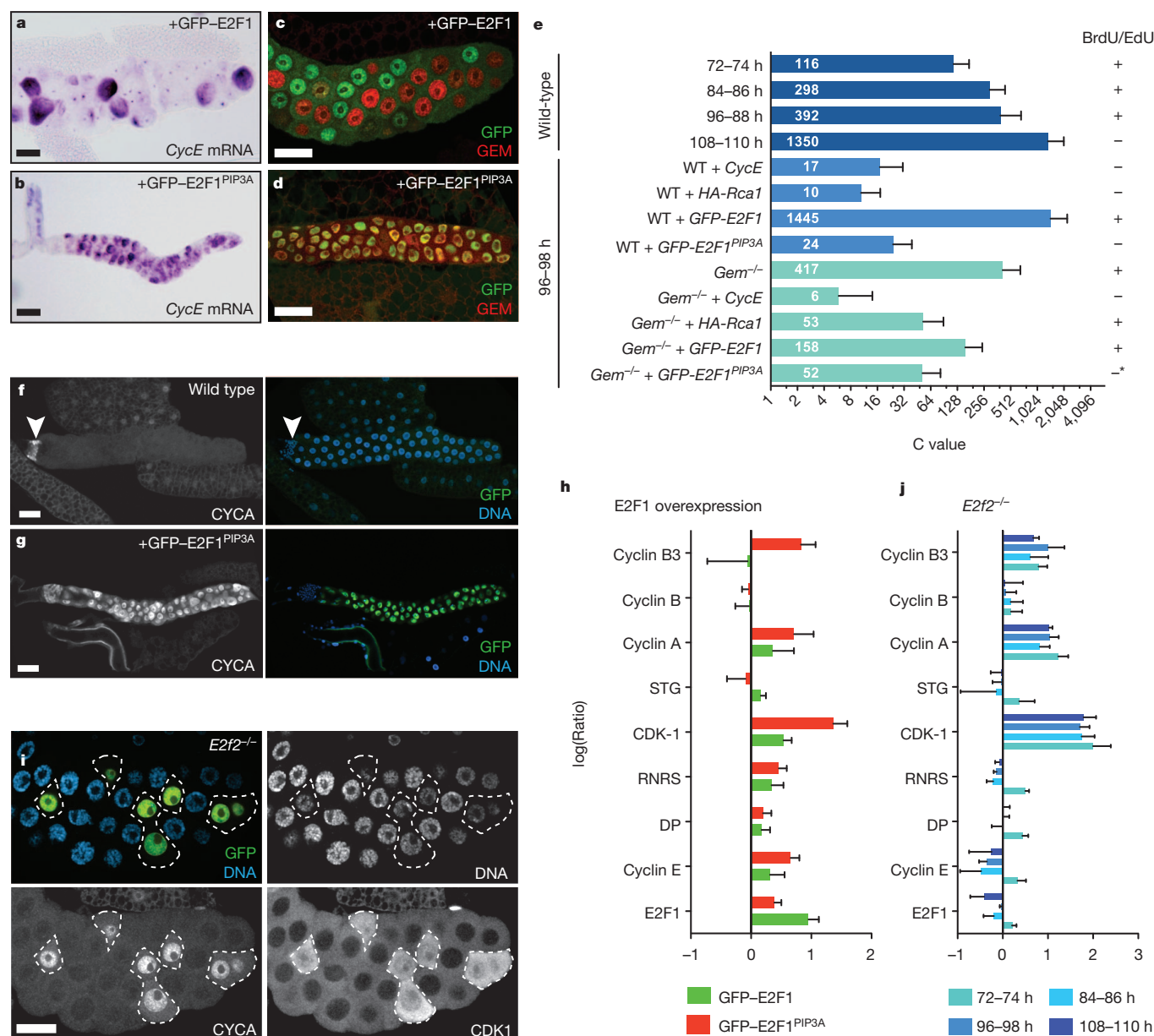


Figure 3 | Endocycle arrest by stabilized E2F1. **a–d**, Expression of WT GFP-E2F1 with *ptc-Gal4* promoted endocycling with cyclic *cycE* (**a**) and *Gem* (**c**), whereas GFP-E2F1^{PIP3A} caused endocycle arrest with uniform *cycE* (**b**) and *Gem* (**d**) expression. **e**, Chromatin (C) values per nucleus for the indicated genotypes and time points. For each genotype about 40 nuclei from 10–20 salivary glands were analysed. Error bars represent standard deviations. *Residual EdU incorporation was detectable, but no endocycle oscillations. **f–g**, CycA expression in WT (**f**) or GFP-E2F1^{PIP3A} (**g**). Arrowhead (**f**) indicates diploid imaginal ring cells. **h**, qRT-PCR measurements of the indicated

mRNAs, from 72 h AED salivary glands expressing GFP-E2F1 (green) or GFP-E2F1^{PIP3A} (red). **i**, CycA and CDK1 accumulation in *E2f2* mutant cells, generated by MARCM mitotic recombination. GFP in **i** marks mutant cells (outlined). CDK1 in bottom right panel was detected using anti-PSTAIRE antibody. **j**, qRT-PCR measurements of the indicated mRNAs, from *E2f2* mutant glands at the indicated time points. Log₁₀(Ratio)s for **h** and **j** are relative to WT controls. Error bars represent standard deviations derived from three to four biological replicates.

cycled, suggesting that this arrest was due to inappropriately timed expression of E2F1 rather than its excessive accumulation (Fig. 3h and Supplementary Fig. 11). Hence S-phase-dependent degradation of E2F1 is essential for endocycling.

One discrepancy between the data and our model was that whereas the model could not readily predict endocycling without E2F (Supplementary Fig. 10), *Dp* and *E2f1 E2f2* mutants support endocycling^{21,22,26}. Our analysis showed that although DP protein was barely detectable in *Dp* mutant glands (Supplementary Fig. 15) cells in these mutants nevertheless endocycled slowly and sustained periodic expression of *CycE* and *RnrS*, and geminin oscillation (Fig. 2b and Supplementary Figs 14 and 15). One explanation for these apparently discrepant observations is that residual maternal E2F activity persists

in these mutants. Consistent with this possibility we found that GFP-E2F1^{PIP3A} was able to block endocycling in *Dp* mutants (Supplementary Fig. 16). Given this observation, the *Dp* mutant phenotype cannot be construed as confounding the model (see Supplementary Discussion).

We next asked how stabilized E2F1 arrests endocycling. Consistent with model predictions, cells arrested by GFP-E2F1^{PIP3A} or CUL-4-RNAi accumulated *CycE* and geminin (Fig. 3). In these arrested cells, however, Geminin accumulation occurred following rather than before arrest (Supplementary Fig. 12), indicating that it did not initiate the arrest. Interestingly, *geminin*-null mutant glands supported rather normal endocycles (Supplementary Fig. 17), but arrest by RCA1 (refs 10, 14), an APC^{Fzr/Cdh1} inhibitor, was substantially rescued in

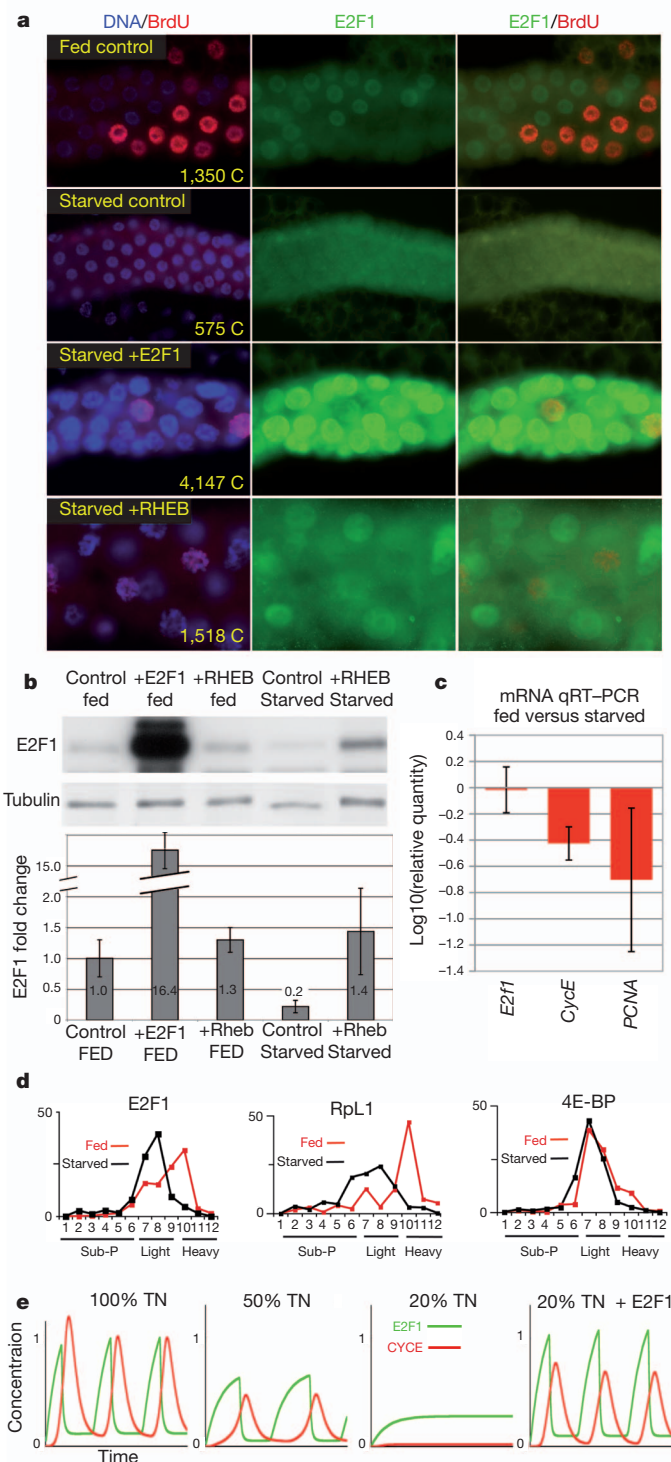


Figure 4 | E2F1 is a growth sensor. **a**, Salivary glands labelled for DNA (blue), E2F1 (green) and incorporated BrdU (red). Fed Control (WT) was labelled with BrdU at 48 h and fixed at 49 h. 'Starved' animals were transferred to protein-free media at 48 h AED, labelled with BrdU at 96 h, and fixed at 97 h AED. *ptc-Gal4* drove expression of *UAS-E2F1/DP* or *UAS-Rheb* in the lower two panels. **c**, mRNA levels from starved and fed control glands, measured by qRT-PCR. **d**, mRNA levels from 3-day protein-starved (black) or fed control (red) whole larvae, quantified from polysome gradient fractions by qRT-PCR. **e**, Computational simulation of starvation by reducing total protein synthesis (TN). In the '20% TN + E2F1' graph, translation of E2F1 was 100% of normal but translation of all other proteins was reduced to 20%. Graphed values (**b**, **c**) include standard deviations calculated from three independent biological samples.

these factors were observed (Fig. 3h), suggesting transcriptional repression. Consistent with this notion these factors were also induced in cells mutant for *E2f2*, *Drosophila's* repressor E2F (Fig. 3i, j). This suggests that, as in mitotic cells²⁷, excess E2F1 may displace E2F2 and thereby derepress its targets. In this context E2F2 appears to act as a selectivity factor that represses mitotic targets in endoreplicating cells. Given that CDK1 is a potent suppressor of PreRCs that can arrest endocycle progression⁶, its derepression probably contributed to endocycle arrest by E2F1^{PIP3A}.

Altogether our results indicate that periodic E2F1 degradation is necessary for endocycling for three reasons: (1) it creates a window of low CYCE/CDK2 activity; (2) it promotes high APC^{Fzr/Cdh1} activity and thereby suppresses geminin accumulation; and (3) it allows E2F2 to maintain repression of CDK1 and its cyclins. Each of these conditions is required for preRC assembly and endocycle progression. This cell cycle mechanism (Fig. 1g and Supplementary Fig. 4) is fundamentally different from that used in mitotic cycles, wherein destruction of the M-phase cyclins by APC^{Cdc20/Fzy}, rather than of E2F1 by the CRL4^{CDT2}, throws the switch that allows preRC assembly⁹. Indeed it is noteworthy that the periodic degradation of E2F1 and depletion of CYCE are not required for mitotic cell cycles in *Drosophila*^{7,12}. CRL4^{CDT2} is required for endocycling in plants⁸, indicating that this element of the endocycle oscillator is conserved.

Finally, we asked what factors control E2F production to regulate endocycle rates. Endocycle speed and number can be manipulated by altering cell growth through changes in dietary protein²⁸ or growth-regulatory genes including *Myc*¹ and insulin/PI3K/TOR signalling components²⁹. Hence we starved larvae of protein to suppress insulin/TOR signalling, reduce protein synthesis, and block cell growth. Starvation arrested the salivary endocycles within 24 h and strongly depleted E2F1 (Fig. 4a, b). *E2f1* and *Dp* mRNA levels were not affected, but the E2F targets *CycE*, *pcna* and *rnrS* were reduced (Fig. 4c, not shown). To test whether this was responsible for starvation-induced endocycle arrest we overexpressed E2F1 in the salivary glands of starved animals. Although these glands failed to grow their nuclei incorporated BrdU and accrued approximately sevenfold more DNA than controls (Fig. 4a). Overexpression of RHEB, which activates the *Target of rapamycin* (TOR) kinase and increases ribosome biogenesis and cap-dependent translation, also restored cell growth, E2F1 protein, and endocycle progression in starved animals (Fig. 4a, b). Thus E2F1 appears to act as a 'growth sensor' that couples rates of endocycle progression to rates of cell growth. A likely mechanism for this, corroborated by modelling (Fig. 4e and Supplementary Fig. 8), involves increased translation of E2F1 in rapidly growing cells. Indeed, we found that the association of E2F1 mRNA with polyribosomes was greatly reduced in protein-starved animals (Fig. 4d). Translational control of E2F is an attractive mechanism for coupling growth to G1/S progression not only in endocycling cells, but also in growth-dependent mitotic cells with extended G1 periods.

the *geminin* mutants (Fig. 3e and Supplementary Fig. 18). This demonstrates that the predominant function of APC^{Fzr/Cdh1} in these endocycles is the degradation of geminin. Importantly, *geminin* mutant cells could be arrested by ectopic CYCE^{10,14} or E2F1^{PIP3A} (Fig. 3e and Supplementary Figs 18, 19). We conclude that whereas geminin accumulation might consolidate the arrest caused by excess E2F1, it is neither initiating nor essential for this arrest.

Further investigations revealed that CYCLIN A, CYCLIN B3 and CDK1 accumulated in E2F1^{PIP3A}-arrested cells (Fig. 3i and Supplementary Fig. 20). These G2/M regulators are not normally expressed in endocycling cells^{8,10}. Large inductions of the mRNAs encoding

METHODS SUMMARY

Larvae were raised at 25 °C in uncrowded conditions, and salivary glands dissected and analysed using standard *Drosophila* genetics and molecular biology methods. DNA quantifications were done using 4',6-diamidino-2-phenylindole (DAPI) fluorescence from CCD images. Computational modelling used delay differential equations tracking the concentrations of mRNAs and proteins, and numerically solved in Mathematica 5.2 (Wolfram Research). Full descriptions of experimental and computational methods, genotypes and reagents is included in the Methods section and Supplementary Information.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 3 August 2010; accepted 21 September 2011.

Published online 30 October 2011.

- Edgar, B. A. & Orr-Weaver, T. L. Endoreplication cell cycles: more for less. *Cell* **105**, 297–306 (2001).
- Lilly, M. A. & Duronio, R. J. New insights into cell cycle control from the *Drosophila* endocycle. *Oncogene* **24**, 2765–2775 (2005).
- Sugimoto-Shirasu, K. & Roberts, K. “Big it up”: endoreduplication and cell-size control in plants. *Curr. Opin. Plant Biol.* **6**, 544–553 (2003).
- Whitman, W. B., Coleman, D. C. & Wiebe, W. J. Prokaryotes: the unseen majority. *Proc. Natl Acad. Sci. USA* **95**, 6578–6583 (1998).
- Follette, P. J., Duronio, R. J. & O’Farrell, P. H. Fluctuations in cyclin E levels are required for multiple rounds of endocycle S phase in *Drosophila*. *Curr. Biol.* **8**, 235–238 (1998).
- Weiss, A., Herzig, A., Jacobs, H. & Lehner, C. F. Continuous cyclin E expression inhibits progression through endoreduplication cycles in *Drosophila*. *Curr. Biol.* **8**, 239–242 (1998).
- Shibutani, S. T. et al. Intrinsic negative cell cycle regulation provided by PIP box- and Cul4^{Cdt2}-mediated destruction of E2f1 during S phase. *Dev. Cell* **15**, 890–900 (2008).
- Roodbarkelari, F. et al. Cullin 4-ring finger-ligase plays a key role in the control of endoreplication cycles in *Arabidopsis* trichomes. *Proc. Natl Acad. Sci. USA* **107**, 15275–15280 (2010).
- Diffley, J. F. Regulation of early events in chromosome replication. *Curr. Biol.* **14**, R778–R786 (2004).
- Zielke, N., Querings, S., Rottig, C., Lehner, C. & Sprenger, F. The anaphase-promoting complex/cyclosome (APC/C) is required for rereplication control in endoreplication cycles. *Genes Dev.* **22**, 1690–1703 (2008).
- Maqbool, S. B. et al. Dampened activity of E2F1-DP and Myb-MuvB transcription factors in *Drosophila* endocycling cells. *J. Cell Sci.* **123**, 4095–4106 (2010).
- Knoblich, J. A. et al. Cyclin E controls S-phase progression and its down-regulation during *Drosophila* embryogenesis is required for the arrest of cell proliferation. *Cell* **77**, 107–120 (1994).
- Lilly, M. A. & Spradling, A. C. The *Drosophila* endocycle is controlled by cyclin E and lacks a checkpoint ensuring S-phase completion. *Genes Dev.* **10**, 2514–2526 (1996).
- Narbonne-Reveau, K. et al. APC/CFzr/Cdh1 promotes cell cycle progression during the *Drosophila* endocycle. *Development* **135**, 1451–1461 (2008).
- Neufeld, T. P., de la Cruz, A. F., Johnston, L. A. & Edgar, B. A. Coordination of growth and cell division in the *Drosophila* wing. *Cell* **93**, 1183–1193 (1998).
- Sigrist, S. J. & Lehner, C. F. *Drosophila* fizzy-related down-regulates mitotic cyclins and is required for cell proliferation arrest and entry into endocycles. *Cell* **90**, 671–681 (1997).
- Shcherbata, H. R., Althausen, C., Findley, S. D. & Ruohola-Baker, H. The mitotic-to-endocycle switch in *Drosophila* follicle cells is executed by Notch-dependent regulation of G1/S, G2/M and M/G1 cell-cycle transitions. *Development* **131**, 3169–3181 (2004).
- Hammond, M. P. & Laird, C. D. Control of DNA replication and spatial distribution of defined DNA sequences in salivary gland cells of *Drosophila melanogaster*. *Chromosoma* **91**, 279–286 (1985).
- Hong, A. et al. The cyclin-dependent kinase inhibitor Dacapo promotes replication licensing during *Drosophila* endocycles. *EMBO J.* **26**, 2071–2082 (2007).
- Duronio, R. J. & O’Farrell, P. H. Developmental control of the G₁ to S transition in *Drosophila*: cyclin E is a limiting downstream target of E2F. *Genes Dev.* **9**, 1456–1468 (1995).
- Royzman, I., Whittaker, A. J. & Orr-Weaver, T. L. Mutations in *Drosophila* DP and E2F distinguish G₁-S progression from an associated transcriptional program. *Genes Dev.* **11**, 1999–2011 (1997).
- Duronio, R. J., Bonnette, P. C. & O’Farrell, P. H. Mutations of the *Drosophila* dDP, dE2F, and cyclin E genes reveal distinct roles for the E2F-DP transcription factor and cyclin E during the S-phase transition. *Mol. Cell. Biol.* **18**, 141–151 (1998).
- Asano, M., Nevins, J. R. & Wharton, R. P. Ectopic E2F expression induces S-phase and apoptosis in *Drosophila* imaginal discs. *Genes Dev.* **10**, 1422–1432 (1996).
- Reis, T. & Edgar, B. A. Negative regulation of dE2F1 by cyclin-dependent kinases controls cell cycle timing. *Cell* **117**, 253–264 (2004).
- Hériché, J. K., Ang, D., Bier, E. & O’Farrell, P. H. Involvement of an SCF^{Slmb} complex in timely elimination of E2F upon initiation of DNA replication in *Drosophila*. *BMC Genet.* **4**, 9 (2003).
- Weng, L., Zhu, C., Xu, J. & Du, W. Critical role of active repression by E2F and Rb proteins in endoreplication during *Drosophila* development. *EMBO J.* **22**, 3865–3875 (2003).
- Frolov, M. V. et al. Functional antagonism between E2F family members. *Genes Dev.* **15**, 2146–2160 (2001).
- Britton, J. S. & Edgar, B. A. Environmental control of the cell cycle in *Drosophila*: nutrition activates mitotic and endoreplicative cells by distinct mechanisms. *Development* **125**, 2149–2158 (1998).
- Britton, J. S., Lockwood, W. K., Li, L., Cohen, S. M. & Edgar, B. A. *Drosophila*’s insulin/PI3-kinase pathway coordinates cellular metabolism with nutritional conditions. *Dev. Cell* **2**, 239–249 (2002).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements Supported by NIH GM51186 to B.A.E., a DAAD fellowship to N.Z., NIGMS 5 P50 GM66050 and NSF MCB0090835 to G.v.D. and K.J.K., DFG LE987/5-1 to C.F.L., CIHR MOP-86622 to S.G., and NIH GM57859 to R.J.D. We thank Y. Liu for help with statistics.

Author Contributions The E2F1-based oscillator was conceived by B.A.E.; N.Z. developed the framework for licensing control and E2F2-mediated repression of mitotic genes. K.J.K. did most of the computational modeling, which was initiated by G.v.D. Initial experiments were done by V.T., who, with help from B.W. and K.J.K., contributed Figs 1d–f, 2a, b, 4a, b and Supplementary Fig. 13. N.Z. carried out much of the later experimental work with help from M.v.S., and contributed Figs 2b, c, 3c–j, 4c and Supplementary Figs 1, 3, 11, 12 and 14–20. S.T.S. and R.J.D. contributed the GFP–E2F1^{PIP3A} transgenics and controls. M.-J.B. contributed Figs 1a–c, 3a, b and Supplementary Fig. 15g, h. S.N. and S.S.G. contributed Fig 4d. C.R. and C.F.L. contributed Supplementary Fig. 2 and the *cdk2*^{−/−} data in Fig. 2b. B.A.E. directed the project and wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to B.A.E. (b.edgar@dkfz.de).

METHODS

Genetics. To express genes in salivary glands *ptc-Gal4*, *43B-Gal4* or *hey-Gal4* females were crossed to males carrying UAS transgenes. *E2f1⁷¹⁷²/E2f1⁷¹⁷²* mutant salivary gland cells were generated by heat-shocking *hs-Flp:FRT82B E2f1⁷¹⁷²/FRT82B ub-GFP-nls* embryos to 37 °C from 2–4 h AED. *Cdk2* mutant glands were generated using the genotype: *F4-Gal4 UAS-GFP/+; Cdk2FRT Cdk2³/Cdk2²* or *F4-Gal4 UAS-GFP/UAS-Flp; Cdk2FRT Cdk2³/Cdk2²*, where *Cdk2FRT* is a transgene encoding an Flp-excisable CDK2.

Mutants:

w; FRT80B, ago¹/TM6B (ref. 30)
y, w, hs-FLP^{1.22}; FRT80B P[mini-w], P[ubi-GFP]/TM6B
dap⁴/CyO, act-GFP (ref. 31)
dap⁸³⁶/CyO, act-GFP (ref. 32)
w; Dp^{a1}/CyO, act-GFP (ref. 21)
w; Dp^{a2}/CyO, act-GFP (ref. 21)
w; FRT42D, Dp^{a3}/CyO, act-GFP (refs 21, 33)
w; Df(2R)Exel7124/CyO, act-GFP (Bloomington *Drosophila* Stock Center number 7872)
w; FRT80B, e2f1⁷¹⁷²/TM6B (*e2f1⁷¹⁷²* is described in ref. 34)
w; e2f2^{76Q1}, cn, bw/CyO, act-GFP (ref. 27)
w; FRT40A, e2f2^{C03344}, dp^{ov}/CyO, act-GFP (gift from M. Frolov)
w; geminin^{(2)k14019}, c, px, sp/CyO, act-GFP (ref. 35)
w; geminin^{(2)k02302}, c, px, sp/CyO, act-GFP (ref. 35)
w; Df(2R)ST1, Adh^{ns}, pr¹, cn/CyO, act-GFP* (ref. 35)

For mutants, we used the strongest alleles available, which in most cases are null alleles. Details on mutant lesions can be found in the cited papers and FlyBase (<http://flybase.org/>).

Transgenes:

ptc-Gal4 (ref. 36)
43B-Gal4 (ref. 5)
w, hey-Gal4, Pin/CyO (Gift from A. Orian)
UAS-Cul4-RNAi (VDR number 44829)
UAS-CycE (ref. 31)
UAS-CycE-RNAi (Nig-Fly number 3938R-3)
UAS-Dap (ref. 31)
UASp-GFP-E2F1 (ref. 7)
UASp-GFP-E2F1-PIP3A (ref. 7)
UAS-E2F1 (ref. 15)
UAS-Rbfl (ref. 37)
UAS-Rheb (ref. 38)
UAS-HA-Rca1 (ref. 39)

Starvation. At 48 h or 72 h AED larvae were washed with PBS and transferred to PBS + 20% sucrose at 25 °C, and maintained on this media until 96 h or 120 h AED, respectively.

DNA quantification. DNA content in nuclei or whole salivary glands was quantified by DAPI fluorescence. Larvae were raised at 25 °C to 96 h AED, and fixed glands were dissected and stained, using an internal control (*ptc-Gal4 UAS-GFPnls*) for each sample. Samples were imaged at ×10 with a CCD camera (Spot RT or Roper HQ2). Average cytoplasmic intensity was subtracted, and the integrated DAPI intensity was used to measure DNA content for whole glands (Fig. 2) or nuclei (Fig. 3). All salivary glands had approximately the same number of cells (<10% variability). Controls were set to 1,350 C according to ref. 18.

Quantification of nuclear concentrations. Nuclear BrdU, Cyclin E, E2F1 and GFP-E2F1 concentrations as shown in Fig. 1d–f and Supplemental Figs 3 and 11 were measured from samples stained with DAPI and the indicated antibodies and imaged by confocal microscopy at ×20. We took image stacks (interval size = 0.65 µm; optimal overlap under our conditions) with optimized imaging conditions such that the deviation from linearity was <10%. To measure average nuclear concentrations of E2F1 and CYCE, we used ImageJ (NIH) and custom software that searched for nuclei by finding ellipsoidal regions that stained brightly for DNA and had the approximate diameter of a nucleus. About half of all nuclei visually overlapped with their neighbours and were not analysed. To reproducibly set the boundaries for each nucleus, we restricted our analysis to optical sections in which the average nuclear DNA staining was >90% maximal (typically 2–5 sections). Mean intensity in these regions was measured in other channels to determine nuclear concentrations.

BrdU labelling. Embryos were collected on grape-juice/agar plates for 2 h and transferred to regular fly food 24 h after egg deposition. At the indicated time points salivary glands were dissected in *Drosophila* Ringer's solution and incubated for 1 h at room temperature with 100 µg ml⁻¹ BrdU in Ringer's solution. Afterwards, the samples were fixed for 30 min in 4% paraformaldehyde/PBS and subsequently treated for 30 min with 2 N HCl. BrdU incorporation was detected with a mouse anti-BrdU antibody (Becton Dickinson) diluted 1:20 in 4% normal goat serum/

PBS, 0.3% Triton-X100 and goat anti-mouse-Alexa Fluor-568 (Invitrogen) as secondary antibody diluted 1:2,000 in 4% NGS/PBS, 0.3% Triton-X100.

EdU labelling. EdU incorporation was performed analogous to the procedure for BrdU labelling using the Click-It EdU Alexa Fluor-555 imaging kit from Invitrogen.

In situ hybridization. Probes for *in situ* hybridization were generated with the DIG RNA labelling system (Roche). For *in vitro* transcription with T7/T3 RNA polymerase the following plasmids were used as template: pT7T3-19U-CycE⁴⁰; pBLu(2)SKM-RnrS⁴¹; pBlusKP-E2F1 (ref. 42). Salivary glands were dissected from larvae staged to the indicated time points. Small batches of about 30 larvae were fixed overnight in 8% formaldehyde/PBS, pooled in scintillation vials and stored until usage in ethanol at –80 °C. The hybridization procedure was performed according to the protocol from ref. 43. For detection samples were probed with the following antibodies: sheep-anti-DIG-AP (1:500, Roche) or mouse anti-DIG-HRP (1:500, Abcam). BCIP/NBT was used as substrate for the AP reaction according to ref. 44, whereas the TSA Alexa Fluor-568 Detection Kit (Invitrogen) was used in combination with horseradish peroxidase.

qRT-PCR. At the indicated time points about 50 salivary glands per genotype were dissected in *Drosophila* Ringer's Solution and immediately transferred to the lysis buffer supplied with the RNeasy mini kit (Qiagen). Samples were stored at –80 °C and then processed with the RNeasy mini kit (Qiagen) according to the manufacturer's instructions including the optional on-column DNase I digestion. 100 ng of total RNA were used for cDNA synthesis with the QuantiTect Reverse Transcription Kit (Qiagen) or the iScript cDNA synthesis kit (Bio-Rad). qRT-PCR data shown in Fig. 3i–k was acquired on a Light Cycler 480 (Roche) using the indicated UPL assays (Roche) and Light Cycler 480 Probes Master (Roche). Relative expression data presented in Fig. 4c was acquired on an iQ5 Instrument (Bio-Rad) using QuantiTect Primer Assays (Qiagen) and the iScript one-step RT-PCR SYBR green kit (Bio-Rad). To ensure statistical significance qRT-PCR was performed in quadruplicates from 3–4 independent samples. Relative expression to GAPDH1 and Actin5c was determined with the $\Delta\Delta C_t$ method:

$$\Delta C_t = C_{t \text{ gene of interest}} - C_{t \text{ endogenous control}}$$

$$\Delta\Delta C_t = \Delta C_t \text{ sample} - \Delta C_t \text{ calibrator}$$

$$\text{Relative quantity} = 2^{-\Delta\Delta C_t}$$

Polysome profiling. Whole larvae were lysed in ice-cold polysome lysis buffer (25 mM Tris pH 6.8, 10 mM MgCl₂, 250 mM NaCl, 1% Triton-X 100, 0.5% sodium deoxycholate, 0.5 µM DTT, 100 µg ml⁻¹ cycloheximide, 10 mg ml⁻¹ heparin, protease inhibitor cocktail (Complete mini, Roche), 2.5 µM PMSE, 5 mM sodium fluoride, 1 mM sodium orthovanadate, RNase inhibitor (RiboLock, Fermentas) using a Dounce homogenizer. Lysates were then cleared by centrifugation (15,000 r.p.m., 15 min, 4 °C in an Eppendorf 5424 centrifuge with an FA-45-24-11 rotor). Equal optical density units (260 nm) of cleared lysates were then layered on 15–45% sucrose gradient (prepared in polysome lysis buffer) and centrifuged (37,000 r.p.m., 2.5 h, 4 °C) in an SW41 Beckman rotor. The gradients were then fractionated using a Brandel BR188 density gradient fractionator with continuous OD (254 nm) reading and collected into twelve equal fractions. The RNA from each fraction was extracted with TRIzol reagent and reversed transcribed using Superscript II (Invitrogen) according to the manufacturer's instructions. Quantitative real-time PCR was then performed as described in ref. 45 using a MyIQ PCR machine (Bio-Rad).

- Moberg, K. H., Mukherjee, A., Veraksa, A., Artavanis-Tsakonas, S. & Hariharan, I. K. The *Drosophila* F box protein archipelago regulates dMyc protein levels *in vivo*. *Curr. Biol.* **14**, 965–974 (2004).
- Lane, M. E. *et al.* Dacapo, a cyclin-dependent kinase inhibitor, stops cell proliferation during *Drosophila* development. *Cell* **87**, 1225–1235 (1996).
- Lane, M. E. *et al.* A screen for modifiers of cyclin E function in *Drosophila melanogaster* identifies Cdk2 mutations, revealing the insignificance of putative phosphorylation sites in Cdk2. *Genetics* **155**, 233–244 (2000).
- Frolov, M. V., Moon, N. S. & Dyson, N. J. dDP is needed for normal cell proliferation. *Mol. Cell. Biol.* **25**, 3027–3039 (2005).
- Duronio, R. J., O'Farrell, P. H., Xie, J.-E., Brook, A. & Dyson, N. The transcription factor E2F is required for S phase during *Drosophila* embryogenesis. *Genes Dev.* **9**, 1445–1455 (1995).
- Quinn, L. M., Herr, A., McGarry, T. J. & Richardson, H. The *Drosophila* Geminin homolog: roles for Geminin in limiting DNA replication, in anaphase and in neurogenesis. *Genes Dev.* **15**, 2741–2754 (2001).
- Speicher, S. A., Thomas, U., Hinz, U. & Knust, E. The *Serrate* locus of *Drosophila* and its role in morphogenesis of the wing imaginal discs: control of cell proliferation. *Development* **120**, 535–544 (1994).
- Xin, S., Weng, L., Xu, J. & Du, W. The role of RBF in developmentally regulated cell proliferation in the eye disc and in cyclin D/Cdk4 induced cellular growth. *Development* **129**, 1345–1356 (2002).
- Saucedo, L. J. *et al.* Rheb promotes cell growth as a component of the insulin/TOR signalling network. *Nature Cell Biol.* **5**, 566–571 (2003).
- Grosskortenhaus, R. & Sprenger, F. Rca1 inhibits APC-Cdh1(Fzr) and is required to prevent cyclin degradation in G2. *Dev. Cell* **2**, 29–40 (2002).

40. Richardson, H. E., O'Keefe, L. V., Reed, S. I. & Saint, R. A. *Drosophila* G1-specific cyclin E homolog exhibits different modes of expression during embryogenesis. *Development* **119**, 673–690 (1993).
41. Duronio, R. J. & O'Farrell, P. Developmental control of a G1-S transcriptional program in *Drosophila*. *Development* **120**, 1503–1515 (1994).
42. Dynlacht, B. D., Brook, A., Dembski, M., Yenush, L. & Dyson, N. DNA-binding and trans-activation properties of *Drosophila* E2F and DP proteins. *Proc. Natl Acad. Sci. USA* **91**, 6359–6363 (1994).
43. Kosman, D. *et al.* Multiplex detection of RNA expression in *Drosophila* embryos. *Science* **305**, 846 (2004).
44. Tautz, D. & Pfeifle, C. A non-radioactive in situ hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene *hunchback*. *Chromosoma* **98**, 81–85 (1989).
45. Van Gilst, M. R., Hadjivassiliou, H. & Yamamoto, K. R. A *Caenorhabditis elegans* nutrient response system partially dependent on nuclear receptor NHR-49. *Proc. Natl Acad. Sci. USA* **102**, 13496–13501 (2005).

Cascades of multisite phosphorylation control Sic1 destruction at the onset of S phase

Mardo Kõivomägi¹, Ervin Valk¹, Rainis Venta¹, Anna Iofik¹, Martin Lepiku¹, Eva Rose M. Balog², Seth M. Rubin², David O. Morgan³ & Mart Loog¹

Multisite phosphorylation of proteins has been proposed to transform a graded protein kinase signal into an ultrasensitive switch-like response^{1–4}. Although many multiphosphorylated targets have been identified, the dynamics and sequence of individual phosphorylation events within the multisite phosphorylation process have never been thoroughly studied. In *Saccharomyces cerevisiae*, the initiation of S phase is thought to be governed by complexes of Cdk1 and Cln cyclins that phosphorylate six or more sites on the Clb5–Cdk1 inhibitor Sic1, directing it to SCF-mediated destruction^{1,5–8}. The resulting Sic1-free Clb5–Cdk1 complex triggers S phase⁹. Here, we demonstrate that Sic1 destruction depends on a more complex process in which both Cln2–Cdk1 and Clb5–Cdk1 act in processive multiphosphorylation cascades leading to the phosphorylation of a small number of specific phosphodegrons. The routes of these phosphorylation cascades are shaped by precisely oriented docking interactions mediated by cyclin-specific docking motifs in Sic1 and by Cks1, the phospho-adaptor subunit of Cdk1. Our results indicate that Clb5–Cdk1-dependent phosphorylation generates positive feedback that is required for switch-like Sic1 destruction. Our evidence for a docking network within clusters of phosphorylation sites uncovers a new level of complexity in Cdk1-dependent regulation of cell cycle transitions, and has general implications for the regulation of cellular processes by multisite phosphorylation.

To study the multiphosphorylation of Sic1, we used a non-inhibitory truncated version of Sic1 (Sic1ΔC) as a substrate for purified Cln2–Cdk1 and Clb5–Cdk1 complexes (Supplementary Fig. 1a; Cdk1 is also known as Cdc28). Intriguingly, both Cln2–Cdk1 and Clb5–Cdk1 generated phosphorylation patterns with abruptly accumulating hyperphosphorylated species (Fig. 1a, b, d, e and Supplementary Fig. 1b). This pattern depended on Cks1, the Cdk1 subunit that binds phosphate groups¹⁰. Mutation of the phosphate-binding site of Cks1 reduced the accumulation of multiphosphorylated forms (Fig. 1a and Supplementary Fig. 2a). Similarly, a phosphorylated competitor phosphopeptide reduced phosphorylation (Fig. 1b). Cks1 mutation and the peptide had little effect on the phosphorylation of a Sic1ΔC version containing a single Cdk site (Fig. 1c).

Cks1 is essential for Cdk1 function^{11,12}, with roles at the G₁/S and G₂/M transitions^{13,14}. We found that the Cks1:Cdk1 stoichiometry *in vivo* was about 1:1 for Cln2–Cdk1 and at least 0.5:1 for Clb5–Cdk1, confirming that Cks1-dependent multiphosphorylation is the prevalent mode of Cdk1 action *in vivo* (Supplementary Fig. 1a–c). An isothermal calorimetry binding assay of fully phosphorylated Sic1ΔC (pSic1ΔC) and Cks1 measured a K_D of $11 \pm 2 \mu\text{M}$, whereas the non-phosphorylated version showed no detectable binding (Supplementary Fig. 1d). Approximately three to four molecules of Cks1 bound each molecule of pSic1ΔC, suggesting that several phosphorylated sites can bind Cks1 independently. Finally, we found that the phospho-binding pocket of Cks1 is required for phosphorylation and degradation of Sic1 *in vivo* (Supplementary Fig. 1e, f).

To understand the Cks1-dependent mechanism, we analysed Sic1ΔC multiphosphorylation over time (Fig. 1d, e). We did not observe significant accumulation of intermediate phosphorylated forms, indicating that phosphorylation was processive. When we performed kinase reactions at Sic1ΔC concentrations higher than apparent K_M , multiphosphorylation patterns remained constant despite the increase in the inhibition term $1 + [S_{0P}]/K_M$ (where S_{0P} is unphosphorylated substrate) (Supplementary Fig. 2b–e). Thus, the enzyme displays processivity; that is, it is able to transfer two or more phosphates to the substrate during a single association event. This conclusion was additionally confirmed using different enzyme concentrations

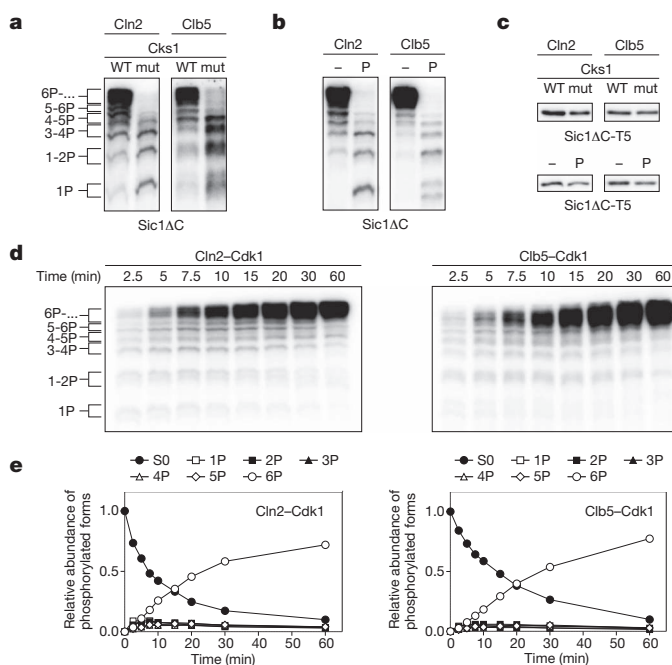


Figure 1 | The phospho-adaptor subunit Cks1 provides processivity for the multiphosphorylation of Sic1 by Cln2–Cdk1 and Clb5–Cdk1. **a**, Cln2–Cdk1 and Clb5–Cdk1 complexes were incubated with Sic1ΔC and [γ -³²P]ATP. The reactions also included wild-type Cks1 (WT) or a version with a mutated phosphate-binding site (mut; see Supplementary Methods). Phosphorylated substrates were separated using Phos-tag SDS–PAGE gels. **b**, Reactions were performed in the presence of a phosphopeptide competitor (P) based on the sequence surrounding T45 in Sic1. **c**, The phosphorylation of a Sic1ΔC version containing a single Cdk site (Sic1ΔC-T5, with other Cdk consensus sites mutated to alanines) was not affected by the Cks1 mutant (mut) or the phosphopeptide. Standard SDS–PAGE was used. **d**, Time courses of Sic1ΔC multiphosphorylation were followed by Phos-tag SDS–PAGE. **e**, The quantified data from **d**. The intensities of ³²P-labelled proteins were divided by the number of phosphates as indicated to obtain the levels of different phosphoforms (see also Methods). In the experiments presented in Fig. 1, enzyme concentrations were chosen to obtain roughly equal substrate labelling.

¹Institute of Technology, University of Tartu, Tartu 50411, Estonia. ²Department of Chemistry and Biochemistry, University of California, Santa Cruz, California 95064, USA. ³Department of Physiology, University of California, San Francisco, California 94158, USA.

in the assay (Supplementary Fig. 2f) and in mathematical simulations (Supplementary Fig. 3). This processive pattern argues against the current model of ultrasensitivity in the Sic1 phosphorylation switch, which is based on the assumption of a distributive mechanism with equal specificity of different sites^{1,15}.

To dissect the mechanism of the processive multiphosphorylation cascade, we first studied the impact of potential docking interactions between Sic1 and cyclins. In previous studies, we found that rapid Sic1 phosphorylation by Clb5–Cdk1 depends on an interaction between RXL motifs in Sic1 and the hydrophobic patch docking site (hp) in Clb5; a triple mutation in this site (Clb5hpm) decreases the net phosphorylation rate (Supplementary Fig. 4a–h)¹⁶. Further, we found here that a version of Sic1ΔC with mutations at its four RXL motifs (Sic1ΔC-1234rxl) showed less abrupt production of multiphosphorylated species by Clb5–Cdk1, showing that processive multiphosphorylation requires both Cks1-dependent and hp-dependent docking (Supplementary Fig. 4i). Cln2–Cdk1 had only a mild RXL effect on the phosphorylation pattern, probably because Cln2 does not contain a conventional hp like that in the B-type cyclins. In recent studies, we also located a ten-amino-acid stretch in Sic1, 136VLLPPSRPTS145, which confers Cln2 specificity¹⁶. Here we found that a five-alanine mutation of the first five hydrophobic residues in this stretch, or a synthetic competitor peptide containing the docking site, reduced the abrupt multiphosphorylation pattern for Cln2 (Supplementary Fig. 4j). In conclusion, both Clb5–Cdk1 and Cln2–Cdk1 use docking mechanisms, in addition to Cks1, to achieve processive multiphosphorylation of Sic1.

Using Sic1ΔC mutants carrying only one Cdk site (Fig. 2a, b), we found that Clb5–Cdk1 rapidly phosphorylated just four sites (T5, T33,

S76 and S80), and this specificity depended on the RXL2 and RXL3 docking sites in Sic1 (Fig. 2b; note that in all figure labels, the indicated Cdk sites are those left unmutated, unless otherwise indicated). Cln2–Cdk1, on the other hand, showed a preference for the N-terminally located site T5 (Fig. 2b). Thus, docking interactions direct the associated kinase to a small number of primary phosphorylation sites. We speculate that these primary sites interact with Cks1 to drive processive phosphorylation of additional sites.

With these primary specificities in mind, we set out to map the pathways along which Cln2–Cdk1 and Clb5–Cdk1 catalyse the phosphorylation of the critical sites required for Sic1 degradation. The original model of Sic1 regulation proposed that six or more sites must be simultaneously phosphorylated *in vivo* to facilitate binding of phospho-Sic1 to the SCF subunit Cdc4 (ref. 1). On the other hand, later binding studies revealed that closely positioned pairs of phosphorylation sites (pT5/pS9, pT45/pT48, or pS76/pS80; see Fig. 2a) each present separate entities with a strong affinity for Cdc4, indicating that just two phosphorylation sites, in the right positions, might be sufficient for Sic1 degradation¹⁷. Our results provided a way to reconcile these findings: we proposed that the requirement for six or more sites *in vivo* reflects a requirement for priming phosphorylation events that direct processive phosphorylation of critical phosphodegrons. To test this possibility, we first measured phosphorylation of a Sic1ΔC mutant with all Cdk sites changed to alanine except for the triple cluster S69/S76/S80, which contains two potential paired degrons (S69/S76 and S76/S80). There was no processive multiphosphorylation of the cluster S69/S76/S80 (Fig. 2c, lane 2 in each panel), but processivity could be induced by adding back single Cdk1 sites to

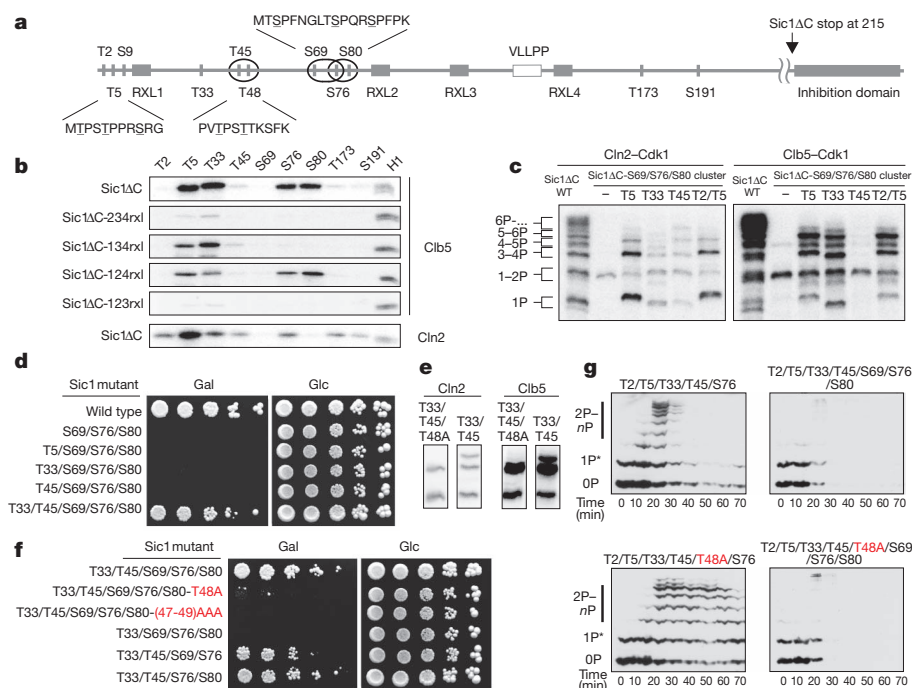


Figure 2 | Phosphorylated priming sites provide docking interactions for efficient phosphorylation of suboptimal sites in phosphodegrons.

a, Schematic view of phosphorylation sites, docking motifs (Clb5- and Cln2-specific), phosphodegrons (ovals¹⁷) in Sic1. **b**, Phosphorylation specificity of Clb5–Cdk1 and Cln2–Cdk1 towards different Cdk sites was studied using Sic1ΔC constructs containing a single fixed Cdk site. For Clb5–Cdk1, the dependence of the site-specificity profile on RXL docking sites was assessed using Sic1ΔC constructs containing a single Cdk site and a single fixed RXL motif. **c**, The impact of different priming phosphorylation sites on cooperative phosphorylation of the degron cluster S69/S76/S80. Phospho-site mutants of Sic1ΔC carrying the intact S69/S76/S80 cluster and the indicated sites left unmutated were used in a kinase assay with Cln2–Cdk1 and Clb5–Cdk1 using Phos-tag SDS–PAGE. **d**, Full-length Sic1 versions containing the combination

of sites described in **c** were overexpressed under the galactose promoter to assay the ability of cells to degrade Sic1. Gal, galactose; Glc, glucose. **e**, Comparison of the *in vitro* phosphorylation profiles of Sic1ΔC versions containing only the phosphorylation sites T33/T45 or T33/T45 with mutation T48A. **f**, The nonconsensus Cdk1 site T48 is important for viability of cells overexpressing Sic1. The same assay as **d** was used. In panels **d** and **f**, the labels indicate unmutated amino acids, and all other consensus Cdk sites are mutated; mutations in the nonconsensus Cdk sites are highlighted in red. **g**, The phosphorylation and degradation dynamics of Sic1 were followed after the release of cells from α -factor in a system constitutively expressing mutated versions of noninhibitory haemagglutinin-tagged Sic1ΔC–3HA. The asterisk indicates a G₁-specific phosphorylation by an unknown kinase.

the amino-terminal side of the cluster. The rate of the appearance of multiphosphorylated species correlated with the site specificity data for Cln2 and Clb5 (Fig. 2b). Addition of the most Cln2-specific site, T5, caused a much greater effect in Cln2 reactions than the less Cln2-specific sites, T33 and T45. Addition of the Clb5-specific sites T5 and T33, but not the poor Clb5 site T45, greatly increased processivity in Clb5 reactions. These results indicate that sites T5, T33 and T45 are able to serve as priming sites for Cks1-dependent phosphorylation of the S69/S76/S80 degon cluster. The phosphopeptide-dependence was confirmed for these mutants as described for Sic1 Δ C-wild type in Fig. 1b (data not shown).

Cells overexpressing Sic1 containing only the triple cluster S69/S76/S80 were inviable (Fig. 2d). Addition of T5, T33 or T45 did not prevent this lethal effect, indicating that phosphorylation of the S69/S76/S80 cluster alone is not sufficient for degradation of Sic1. However, viability improved when both T33 and T45 were added. Notably, the addition of T45, together with a nonconsensus Cdk site T48 (Fig. 2a), creates a double degon, T45/T48, as predicted previously¹⁷. T48 is known to be phosphorylated *in vivo*⁷. We proposed that T33 serves as a docking site for both the T45/T48 and S69/S76/S80 clusters, and that T45 serves both as a constituent site of the T45/T48 degon as well as a Cks1-dependent docking site for the degon cluster S69/S76/S80. Indeed, T48 alone (the Sic1-9A mutant with all Cdk sites changed to alanine) was a very poor substrate for Clb5 and Cln2 (data not shown), but the addition of T33 or the T33/T45 pair made it a specific site (Fig. 2e, Supplementary Fig. 5 and Supplementary Table 1), implying that T33 phosphorylation allows bypass of the consensus motif requirement of a +1 proline at T48. Notably, we found that the alanine mutation in T48 had a strong growth-suppressing effect in the galactose assay within the context of sites T33/T45/S69/S76/S80 and also had a weaker effect in the background containing all the Cdk sites (Fig. 2f and Supplementary Fig. 4k). Our results indicate that the benefit of multisite phosphorylation of Sic1, compared to a system with a single phosphorylated site with high affinity for Cdc4 (for example, based on an optimal degon site of cyclin E¹), is likely to be the ability of this mechanism to provide

docking-dependent kinase specificity for paired degons. Thus, phosphorylation sites in Sic1 can be divided into three categories: (1) paired degon sites that are spaced 3–7 amino acids from each another; (2) N-terminally positioned priming sites for each paired degon; and (3) sites that serve as both priming and degon sites (for example, T45).

It is unclear why a single paired degon with a priming site is insufficient for degradation of Sic1, as suggested by our observation that viability in the galactose plate assay requires both T45/T48 and either S69/S76 or S76/S80 (Fig. 2f). To test if the simultaneous presence of both paired degons is required for degradation of Sic1, we used western blotting to measure phosphorylation and degradation of mutated versions of Sic1 Δ C. Remarkably, a Sic1 Δ C construct (equivalent to the Sic1-5p mutant of ref. 1) containing only the degon T45/T48, but missing the intact degon around S76, was rapidly degraded after the release of cells from G₁ (Fig. 2g). This degradation was abolished by mutation of the single nonconsensus Cdk site at T48. A construct also containing the S69/S76/S80 degon (Sic1-7p of ref. 1) was more rapidly degraded, and mutation of T48 in this background did not influence the degradation rate. We conclude that the T45/T48 degon is sufficient to promote some degradation of Sic1 *in vivo*, but this rate of degradation is not sufficient to prevent the lethality of overexpressed Sic1.

Our model assumes differential roles of Cln2 and Clb5 in the order of Sic1 phosphorylation events. To explore this possibility, we developed methods for determining the apparent rate constants, which we termed k_{dock} , for individual Cks1-enhanced phosphorylation steps (Fig. 3a, b and Supplementary Table 1). The results revealed considerable differences between Clb5 and Cln2. Clb5–Cdk1 was much more effective than Cln2–Cdk1 in taking shortcuts to the critical degon pair of S76/S80, using T5 and T33 as priming sites for Cks1, and with assistance from RXL-mediated docking (Supplementary Tables 1 and 2, Fig. 3a–c and Supplementary Fig. 6). Notably, in the case of Clb5, different RXL motifs supported different Cks1-dependent docking events (Supplementary Table 2).

We propose that in late G₁, Clb5–Cdk1 is inhibited by Sic1, and the cascade of phosphorylation events begins with T5 phosphorylation by

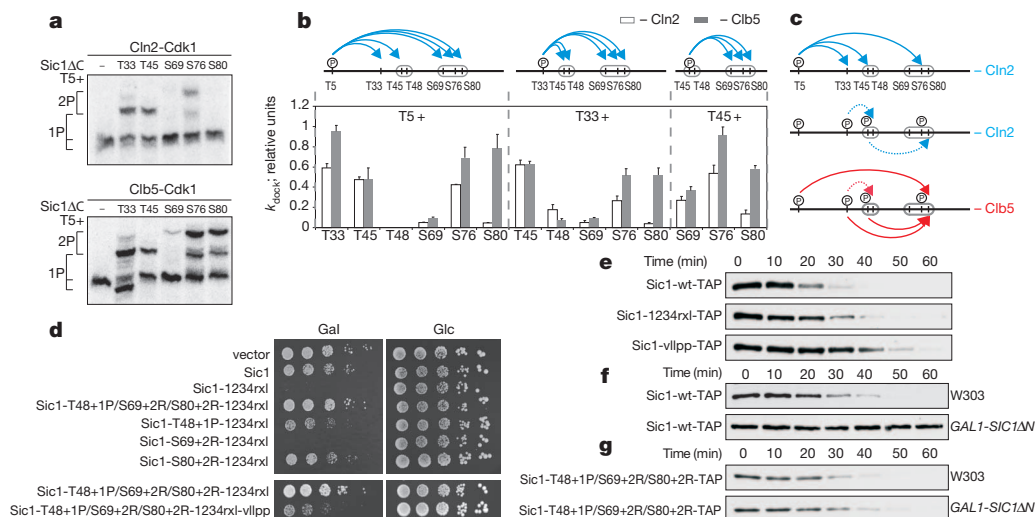


Figure 3 | Differential roles of Cln2 and Clb5 in Sic1 multiphosphorylation and degradation. **a**, Pair-wise mapping of the docking connections underlying Sic1 multiphosphorylation, using purified Sic1 Δ C mutants containing just two of the Cdk phosphorylation sites per mutant. Representative examples of autoradiographs of phosphorylation assays, showing different docking specificities between Cln2–Cdk1 and Clb5–Cdk1. **b**, The specificity profiles for different pair-wise docking connections. The error bars indicate standard errors of the means of at least two independent experiments. (See Supplementary Information and Supplementary Table 1). **c**, Schematic view of fast and slow docking-dependent phosphorylation steps for Cln2–Cdk1 and Clb5–Cdk1. **d**, Sic1 mutants with improved Cdk recognition determinants in suboptimal phosphodegrons rescue the inviability of cells overexpressing Sic1-

1234rxl. The Cln2-dependent docking site becomes essential under these conditions. **e**, Cells carrying *SIC1wt-TAP* or versions with docking site mutations at the endogenous *SIC1* locus were released from an α -factor arrest and the degradation pattern of Sic1–TAP protein was followed by western blotting, using standard SDS–PAGE. **f**, A wild-type strain (*SIC1wt-TAP*) and a strain also expressing the nondegradable inhibitory domain of Sic1 (*SIC1 Δ N*) under the *GAL1* promoter were arrested in α -factor, followed by the addition of galactose. After 45 min, the cells were washed into galactose media lacking α -factor and Sic1–TAP levels were followed by western blotting (see also Supplementary Fig. 7). **g**, Sic1 mutants with improved Cln–Cdk recognition determinants in suboptimal phosphodegrons trigger rapid degradation of Sic1 in the presence of *SIC1- Δ N*. Experiment was performed as in panel **f**.

Cln2–Cdk1. This priming event is followed by docking-enhanced phosphorylations leading to a phosphorylated chain of sites pT5/pT33/pT45/pS76 but not fully phosphorylated paired degrons, as phosphorylation by Cln2–Cdk1 of suboptimal sites in the degrons (T48 and S69, or S80) is slow (Fig. 3a, b). However, the phosphorylated cluster pT5/pT33/pT45/pS76 serves as a powerful Cln2–Cdk1-dependent docking platform for emerging Clb5–Cdk1. As Cln2 levels rise, such priming forces would create a synergistic effect between Cln2 and Clb5, greatly amplifying the impact of low emerging levels of free Clb5–Cdk1 complexes and defining the point of no return for Clb5-dependent positive feedback. A prediction of this model is that changing the limiting suboptimal degron sites to optimal Cdk sites will rescue the lethality of Sic1-1234xrl (Supplementary Fig. 4f), as the degradation in this case should be driven primarily by Cln2. Indeed, changing T48 to a Cdk1 site by introducing a proline at position +1 partially rescued the lethal phenotype of Sic1-1234xrl (Fig. 3d). A similar rescue was attained by introducing a positive determinant for Cln2–Cdk1, an arginine at position +2 (ref. 16), to the site S80. Optimization of the S69 site had no effect. Almost complete rescue was gained by a triple mutation with all three limiting degron sites (T48, S69 and S80) changed to optimal Cdk sites. Importantly, these effects are unlikely to be due to improved binding of phosphodegrons to Cdc4, as the basic residues on the carboxy-terminal side of pS/pT are known to disrupt the Cdc4 interaction¹. Finally, to confirm that degradation of these Sic1 mutants is driven by Cln–Cdk1, instead of Clb5–Cdk1, we additionally mutated the Cln2-specific docking site VLLPP in the triple mutant background (Fig. 3d, lower panel). The *vllpp* mutation abolished the rescue effect of the triple mutant. These data indicate that Cln2 alone does have the potential to drive Sic1 degradation, but the Cln2-driven phosphorylation cascade is terminated at the rate-limiting final steps. However, this mechanism allows creation of the Clb5 docking platform containing the chain of optimal sites pT5/pT33/pT45/pS76.

Finally, to compare further the functions of Clb5 and Cln2, we analysed the degradation of endogenous Sic1. We found that mutation of either the Cln-specific *vllpp* docking motif or Clb5-specific RXL docking sites delayed Sic1 degradation (Fig. 3e), confirming that both Cln2 and Clb5 have a role in the timing of Sic1 degradation. However, when all Clb–Cdk1 activity in the cell was specifically inhibited by overexpression of nondegradable Sic1, endogenous Sic1 was completely stabilized (Fig. 3f and Supplementary Fig. 7), indicating that the key trigger for Sic1 degradation and the G₁/S transition is the emerging free Clb5–Cdk1, after its levels exceed those of the inhibitory complex.

The inability of Cln alone to cause Sic1 degradation could be attributed to the slow phosphorylation rate of sites in the S76/S80 degron. Indeed, by introducing optimal Cln consensus motifs into the slow degron sites, analogously to the experiment described in Fig. 3d, the Sic1 degradation pattern was restored to normal despite the absence of Clb–Cdk1 activity (Fig. 3g). The processive multiphosphorylation cascades, composed of a set of fast and slow steps and different docking specificities, enable this discrimination between the signal outputs of different cyclin–Cdk1 complexes. Furthermore, the Cln output state acts as a primer state for the second complex, creating a potential ‘AND gate’ in which Cln2 is not allowed to trigger the G₁/S transition until sufficient levels of Clb5 activity accumulate.

In conclusion, our model provides novel insights into the multisite phosphorylation mechanism of Sic1 and, potentially, of Cdk1 targets in general. The multiple sites create a network of docking connections that exploit Cks1-dependent and cyclin-specific docking interactions to process Cdk1 signals to achieve proper tuning of the timing of the G₁/S transition (Supplementary Fig. 8). As most Cdk1 targets in the cell contain clusters of multiple sites¹⁸, the regulation of cell cycle switchpoints by Cdk1-dependent multiphosphorylation might prove to be far more complex than generally anticipated, and it is possible

that beneath the seemingly random constellations of phosphorylation sites, an intricate signal processing logic may be hidden.

METHODS SUMMARY

Yeast strains were in the W303 background and are listed in Supplementary Table 3. Plasmid constructs are listed in Supplementary Table 4. Phos-tag Acrylamide AAL-107 (ref. 19) was purchased from NARD Institute.

Full Methods and any associated references are available in the online version of the paper at www.nature.com/nature.

Received 14 February; accepted 20 September 2011.

Published online 12 October 2011.

- Nash, P. *et al.* Multisite phosphorylation of a CDK inhibitor sets a threshold for the onset of DNA replication. *Nature* **414**, 514–521 (2001).
- Borg, M. *et al.* Polyelectrostatic interactions of disordered ligands suggest a physical basis for ultrasensitivity. *Proc. Natl Acad. Sci. USA* **104**, 9650–9655 (2007).
- Kim, S. Y. & Ferrell, J. E. Jr. Substrate competition as a source of ultrasensitivity in the inactivation of Wee1. *Cell* **128**, 1133–1145 (2007).
- Thomson, M. & Gunawardena, J. Unlimited multistability in multisite phosphorylation systems. *Nature* **460**, 274–277 (2009).
- Tang, X. *et al.* Suprafacial orientation of the SCF^{Cdk1} dimer accommodates multiple geometries for substrate ubiquitination. *Cell* **129**, 1165–1176 (2007).
- Tyers, M. The cyclin-dependent kinase inhibitor p40^{Sic1} imposes the requirement for Cln G1 cyclin function at Start. *Proc. Natl Acad. Sci. USA* **93**, 7772–7776 (1996).
- Verma, R. *et al.* Phosphorylation of Sic1p by G₁ Cdk required for its degradation and entry into S phase. *Science* **278**, 455–460 (1997).
- Cross, F. R., Schroeder, L. & Bean, J. M. Phosphorylation of the Sic1 inhibitor of B-type cyclins in *Saccharomyces cerevisiae* is not essential but contributes to cell cycle robustness. *Genetics* **176**, 1541–1555 (2007).
- Schneider, B. L., Yang, Q. H. & Futcher, A. B. Linkage of replication to start by the Cdk inhibitor Sic1. *Science* **272**, 560–562 (1996).
- Arvai, A. S., Bourne, Y., Hickey, M. J. & Tainer, J. A. Crystal structure of the human cell cycle protein CksHs1: single domain fold with similarity to kinase N-lobe domain. *J. Mol. Biol.* **249**, 835–842 (1995).
- Hadwiger, J. A., Wittenberg, C., Mendenhall, M. D. & Reed, S. I. The *Saccharomyces cerevisiae* Cks1 gene, a homolog of the *Schizosaccharomyces pombe* *suc1*⁺ gene, encodes a subunit of the Cdc28 protein kinase complex. *Mol. Cell. Biol.* **9**, 2034–2041 (1989).
- Reynard, G. J., Reynolds, W., Verma, R. & Deshaies, R. J. Cks1 is required for G₁ cyclin-cyclin-dependent kinase activity in budding yeast. *Mol. Cell. Biol.* **20**, 5858–5864 (2000).
- Tang, Y. & Reed, S. I. The Cdk-associated protein Cks1 functions both in G₁ and G₂ in *Saccharomyces cerevisiae*. *Genes Dev.* **7**, 822–832 (1993).
- Patra, D., Wang, S. X., Kumagai, A. & Dunphy, W. G. The *Xenopus* Sic1/Cks protein promotes the phosphorylation of G₂/M regulators. *J. Biol. Chem.* **274**, 36839–36842 (1999).
- Deshaies, R. J. & Ferrell, J. E. Jr. Multisite phosphorylation and the countdown to S phase. *Cell* **107**, 819–822 (2001).
- Köivomägi, M. *et al.* Dynamics of Cdk1 substrate specificity during the cell cycle. *Mol. Cell* **42**, 610–623 (2011).
- Hao, B., Oehlmann, S., Sowa, M. E., Harper, J. W. & Pavletich, N. P. Structure of a Fbw7-Skp1-cyclin E complex: multisite-phosphorylated substrate recognition by SCF ubiquitin ligases. *Mol. Cell* **26**, 131–143 (2007).
- Holt, L. J. *et al.* Global analysis of Cdk1 substrate phosphorylation sites provides insights into evolution. *Science* **325**, 1682–1686 (2009).
- Kinoshita, E., Yamada, A., Takeda, H., Kinoshita-Kikuta, E. & Koike, T. Novel immobilized zinc(II) affinity chromatography for phosphopeptides and phosphorylated proteins. *J. Sep. Sci.* **28**, 155–162 (2005).

Supplementary Information is linked to the online version of the paper at www.nature.com/nature.

Acknowledgements We thank D. Kellogg for strains and L. Peil for advice on mass spectrometry. This work was supported by International Senior Research Fellowship No. 079014/Z/06/Z from the Wellcome Trust (M.Lo.), an installation grant from EMBO and HHMI (M.Lo.), no. 1253, grants no. 6766 from the Estonian Science Foundation (M.Lo.) and SF0180071s07 from Estonian Ministry of Education and Research (M.Lo.), EMP grant no. 08071N from the Norwegian government, and grants from the National Institute of General Medical Sciences (D.O.M.) and National Cancer Institute (S.M.R.).

Author Contributions M.K., E.V., R.V., A.I., M.Le. and M.Lo. designed and performed the experiments, except for the isothermal calorimetry experiments, performed by E.R.M.B. and S.M.R.; M.Lo. coordinated the project and wrote the manuscript with assistance from D.O.M. and S.M.R.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of this article at www.nature.com/nature. Correspondence and requests for materials should be addressed to M.Lo. (Mart.Loog@ut.ee).

METHODS

Protein purification. TAP-purification of cyclin-Cdk1 complexes was performed as described previously^{20,21} using C-terminally TAP-tagged cyclin constructs cloned into 2 micron vectors and overexpressed from the *GAL1* promoter. For purification of 3HA-Clb2-Cdk1, a yeast strain (a gift from D. Kellogg) with the *GAL1* promoter introduced along with the N-terminal 3HA tag in the chromosomal locus of the *CLN2* gene was used. The overexpressed 3HA-Clb2-Cdk1 complex was purified as described²², using immunoaffinity chromatography with a rabbit polyclonal antibody against the haemagglutinin epitope (purchased from Labas). Purification of N-terminally 6His-tagged recombinant Sic1 constructs was performed using cobalt affinity chromatography. Cks1 was purified as described previously¹².

Phosphorylation assays. For quantitative phosphorylation assays, the substrate concentration was kept in the range of 0.5–2 μM (in the linear $[S]$ versus v_0 range, several-fold below estimated K_M) and the initial velocity conditions were defined as a substrate turnover ranging up to 10%/N of the total concentration of N Cdk sites. The general composition of the assay mixture was as follows: 50 mM HEPES pH 7.4, 5 mM MgCl_2 , 150 mM NaCl, 0.1% NP-40, 20 mM imidazole, 2% glycerol, 2 mM EGTA, 0.2 mg ml⁻¹ BSA, 500 nM Cks1 and 500 μM ATP (with added $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ (Perkin Elmer)). Around 1–10 nM of the purified kinase complex was used, the amount depending on the setup of the experiment. The optimal working concentration for purified Cks1 was taken as 500 nM according to the optimization performed for cyclin-Cdk1 preparations using Sic1 ΔC as a substrate. For the phosphorylation assay with mutant Cks1, purified kinase complexes were preincubated for 45 min with Cks1 wild type or mutant to compensate for differences in the amounts of Cks1 already present in the preparations. The composition of the preincubation mixture was: 50 mM HEPES pH 7.4, 5 mM MgCl_2 , 150 mM NaCl, 0.2 mg ml⁻¹ BSA and 500 μM ATP. Kinase assay was initiated by adding preincubation mixture and $[\gamma\text{-}^{32}\text{P}]\text{ATP}$ to the substrate. Aliquots were taken at least at two different time points and the reaction was stopped by SDS-PAGE sample buffer. For separating the phosphorylated versions of Sic1, 10% SDS-PAGE was used, supplemented with the Phos-tag reagent¹⁹ according to the instructions from the manufacturer. For quantitative analysis of the dynamics of multisite phosphorylation (in Fig. 1d, e and Supplementary Fig. 2a–d, f) the six major bands separated in Phos-tag SDS-PAGE were quantified as 1- to 6-phosphate species.

For the phosphorylation assay of full-length Sic1 as part of the inhibitory complex, Clb5-TAP-Cdk1 was isolated from yeast cell extract containing overexpressed Clb5-TAP with IgG beads (Supplementary Fig. 4b). The stoichiometric Sic1-Clb5-Cdk1 complex was formed by incubating an excess amount of purified Sic1 with the beads, and unbound Sic1 was removed by washing (50 mM HEPES pH 7.4, 0.5 M NaCl, 0.1% NP-40). The phosphorylation reaction was performed according to the standard kinase assay protocol and was initiated by adding purified cyclin-Cdk1 complexes to the washed beads. The Clb5-dependent phosphorylation of full-length Sic1 as part of the stoichiometric inhibitory complex revealed a similar RXL specificity profile (Supplementary Fig. 4b). This result also suggests that Sic1 ΔC is a valid model substrate, which was chosen for large scale analysis instead of the hard-to-adjust assay with the stoichiometric complex. It is reasonable to consider the N-terminal region of Sic1 as an independent polypeptide entity, given that only a short C-terminal part is required for high-affinity inhibition and that Sic1 is an intrinsically disordered protein²³.

The mutant Cks1 used in our experiments was designed to disrupt the phosphate-binding site and contained the combination of mutations R33E, S82E and R102A (ref. 24). The triple mutants in the hydrophobic patch (hpm) of Clb5 and Clb2 were described previously²⁵, and the hpm of Clb3 (F201A, L205A, T208A) was designed according to sequence homology with other B-type cyclins.

The relative rate constants for different phospho-docking enhanced steps (k_{dock}) were determined using mutated Sic1 ΔC versions containing different pair-wise combinations of Cdk1 sites (the rest of the serine and threonine residues in S/TP motifs were mutated to alanines). It is impossible to estimate the effects of docking on the reaction rates directly, as it is hard to produce a version of substrate protein where the primed docking sites are fully phosphorylated and the secondary sites are unphosphorylated. Therefore, we used an indirect approach by estimating the relative formation rate of doubly phosphorylated species. The k_{dock} was defined as the ratio of the observed fraction of the doubly phosphorylated form and the estimated kinase activity towards the single N-terminal priming site present in the pair-wise Sic1 ΔC construct. The latter parameters were estimated using quantified values for single site specificities from the experiment in Fig. 2b (in the absence or presence of different docking sites). Thus, the k_{dock} values are independent of the rates of the priming steps and reflect only the rates of the secondary steps. The phosphorylation of the substrate was followed in a conventional kinase assay and singly and doubly phosphorylated species were resolved using Phos-tag SDS-PAGE and quantified by PhosphorImager. The k_{dock} was calculated from two

consecutive time points (that is, 8 and 16 min, in the low initial range of total substrate consumption) from at least two independent experiments. We applied a condition of a minimal ratio of 0.7 for the k_{dock} values from these two time points, to ensure that the singly phosphorylated species had not reached the temporary quasi-steady state. The obtained values are 'apparent constants' as it is impossible to precisely determine the relative contributions of processive and distributive mechanisms. Nevertheless, the obtained k_{dock} values provide very good estimates of how much the phosphorylation of a site is enhanced when another site is present.

Western blotting and viability assays. For viability assay, log-phase cultures were equalized in density and spotted as serial dilutions on selective synthetic complete (SC) plates. The plates were incubated for 48–60 h at 30 °C. For the western blotting experiments, cells were grown to OD_{600 nm} = 0.3 and treated for 2.5 h with 1 $\mu\text{g ml}^{-1}$ α -factor or for 2.5 h with 15 mg ml⁻¹ hydroxyurea and released by washing. In the experiments presented in Fig. 3f, g and Supplementary Fig. 7 the *GAL1-SIC1 ΔN* was integrated into the *URA3* locus. In Fig. 2g the Sic1 ΔC -3HA versions were cloned into the pRS315 vector and constitutively expressed under the *ADH* promoter. In the experiments presented in Fig. 3e, f, g the endogenous *SIC1* was C-terminally TAP-tagged. The cells were lysed by bead-beating in lysis buffer containing urea. Blotting of Phos-tag SDS-PAGE gels was performed using a dry system iBlot (Invitrogen). The antibody used for western blotting of 3HA-tagged proteins was HA.11 clone 16B12 from Covance and the antibodies used for western blotting of Cdk1 (Cdc28 (yC-20)) and of TAP-tagged proteins (c-Myc (A-14)) were from Santa Cruz. The rabbit anti-Cks1 antibody was from Labas.

Isothermal calorimetry. Recombinant Sic1 ΔC was phosphorylated with purified Clb2-Cdk1, and complete phosphorylation was confirmed by following the phosphorylation shift by Phos-tag SDS-PAGE. Recombinant Cks1 was expressed in *Escherichia coli* from a pET vector and purified with anion exchange and size-exclusion chromatography. Calorimetry experiments were performed with a VP-ITC system (MicroCal). 0.3–0.4 mM Cks1 was titrated into a 30 μM solution of phospho-Sic1 ΔC . Experiments were carried out at 25 °C in a buffer containing 25 mM Tris and 150 mM NaCl (pH 8.0). Data were analysed with the MicroCal Origin software package. The reported binding constant and stoichiometry are the average from 2 experiments, and the reported errors are the standard deviation of these measurements.

Quantitative mass spectrometry. For quantitative determination of Cks1-dependent phosphorylation of T48, equal amounts of Sic1 ΔC -wt protein were phosphorylated by Clb5-Cdk1 supplemented with normal isotopic ATP ($[\text{O}]\text{ATP}$) or heavy ATP ($[\text{O}]\text{ATP}$) (Cambridge Isotope Laboratories). Kinase assays were incubated at room temperature for 60 min then pooled together in a 1:1 ratio (v/v) in SDS-PAGE sample buffer. The proteins were separated by 10% SDS-PAGE and the gels were stained with Coomassie brilliant blue G-250 (Sigma) and protein bands were excised from the stained gels. Proteins were in-gel digested by LysC/P (10 ng μl^{-1}) (Wako) and peptides were purified by using C18 StageTips.

Peptides were separated with an Agilent 1200 series nanoflow system (Agilent Technologies) connected to a LTQ Orbitrap classic mass-spectrometer (Thermo Electron) equipped with nanoelectrospray ion source (Proxeon). Purified peptides were loaded on a fused silica emitter (75 $\mu\text{m} \times 150 \text{ mm}$) (Proxeon) packed in-house with Reprosil-Pur C18-AQ 3- μm particles (Dr. Maisch HPLC GmbH). Peptides were separated with 30 min 3–40% B gradient (A, 0.5% acetic acid; B, 0.5% acetic acid/80% acetonitrile) at a flow-rate of 200 nl min⁻¹, eluted peptides were sprayed directly into LTQ Orbitrap mass-spectrometer with a spray voltage of 2.2 kV. The mass spectrometry scan range was m/z 300–1,800 and the top 5 precursor ions were selected for subsequent MS/MS scans. A lock-mass was used for the LTQ-Orbitrap to obtain constant mass-accuracy during the gradient analysis. Peptides were identified with the Mascot 2.3 (<http://www.matrixscience.com>) search engine. A peptide mass tolerance of 7 p.p.m. was used and a fragment ion mass tolerance of 0.6 Da. Two missed cleavage sites for LysC/P were allowed. The oxidation of the methionine and the phosphorylation of serine, threonine were set as variable modifications.

- Puig, O. *et al.* The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* **24**, 218–229 (2001).
- Ubersax, J. A. *et al.* Targets of the cyclin-dependent kinase Cdk1. *Nature* **425**, 859–864 (2003).
- McCusker, D. *et al.* Cdk1 coordinates cell-surface growth with the cell cycle. *Nature Cell Biol.* **9**, 506–515 (2007).
- Mittag, T. *et al.* Structure/function implications in a dynamic complex of the intrinsically disordered Sic1 with the Cdc4 subunit of an SCF ubiquitin ligase. *Structure* **18**, 494–506 (2010).
- Bourne, Y. *et al.* Crystal structure and mutational analysis of the *Saccharomyces cerevisiae* cell cycle regulatory protein Cks1: implications for domain swapping, anion binding and protein interactions. *Structure* **8**, 841–850 (2000).
- Loog, M. & Morgan, D. O. Cyclin specificity in the phosphorylation of cyclin-dependent kinase substrates. *Nature* **434**, 104–108 (2005).

CORRIGENDUM

doi:10.1038/nature10604

DNA demethylation for hormone-induced transcriptional derepression

Mi-sun Kim, Takeshi Kondo, Ichiro Takada, Min-young Youn, Yoko Yamamoto, Sayuri Takahashi, Takahiro Matsumoto, Sally Fujiyama, Yuko Shiode, Ikuko Yamaoka, Hirochika Kitagawa, Ken-ichi Takeyama, Hiroshi Shibuya, Fumiaki Ohtake & Shigeaki Kato

Nature 461, 1007–1012 (2009)

Several lanes of the ChIP analyses in this Letter were inadvertently duplicated or erroneously created during figure assembly. We now provide corrected figure panels for Figs 1f, 2c, 2f, 2g and 3h and Supplementary Figs S8, S9a, S9b, S11, S13b, S18 and S28. Our results and conclusions are not affected by these errors, but we apologise for the careless mistakes made.

During initial preparation of the figure panels, the panels for the negative controls (with no obvious signals), inputs and some data were inappropriately duplicated and mixed up. The experiments were performed several times, so a set of data with adequate negative controls and inputs from one experimental round could be found. We have replaced the previous set of data in these panels with a correct set, and confirmed that these corrections do not affect the original claims in the published text. Representative ChIP data are displayed and significance was statistically assessed from several independently repeated experiments. We have also repeated the experiments and obtained the same results as those in the published figures.

In Fig. 1f, lane 5 of anti-AcH3, lane 1 of anti-5-MetC, and the anti-IgG panel of a previous data set were left blank without placing of gel images. In Fig. 2c, the panels of distal regions (–3,907/–3,442; left panels) and the input panels (–230/+130) were duplicated. The predicted data of anti-5MetC was inadvertently mixed up. In Figs 2f, 2g and 3h, the input panels were duplicated.

Further errors in the Supplementary Figures of the original Letter are described and corrected in the Supplementary Information of this Corrigendum, and the supporting raw data is also presented.

Supplementary Information is linked to the online version of this Corrigendum at www.nature.com/nature.

